VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2021, ISSN (e): 2309-3978, ISSN (p): 2411-6246
Volume 9, Number 4, October-December, 2021 *pp.: 92-101*

Use of big data governance in several corporate sectors

Mansoor Ahmad Rasheed*, Hafiz Zeeshan Ahmad, Hafiz Abdullah Tanweer, Hudabia Murtaza, Mannan Ahmad Rasheed & Mishaal Ahmed

University of Management and Technology, Lahore Pakistan
*Corresponding author email address: S2020114015@umt.edu.pk, mansoorahmadrasheed@gmail.com

ABSTRACT
Big data is rapidly in use by different organizations. It is another step towards technological advancement wsith tremendous benefits to almost all fields. Organizations need big data to find out the needs and priorities of the customers and to make strategies. A large amount of data is called big data. Big data is complex data that is not so easy to manage for that purpose big data governance is developed. It manages big data in all possible ways. The important information required is separated through big data governance techniques. This research is based on the importance of big data governance which is provided by digging out the big data governance frameworks, tools, and techniques for carrying out the data governance. There are many challenges and opportunities related to big data governance that are also mentioned in the paper. A brief discussion on the best practices of big data governance is provided along with a case study of the national pension service of South Korea. The report ends with a detailed conclusion based on the finding of the paper with some future insight.

KEYWORDS
Big data, big data governance, big data governance framework, big data Opportunity & challenges, Big data in different organizations

JOURNAL INFO
HISTORY: Received: October 11, 2021 Accepted: December 15, 2021 Published: December 31, 2021

INTRODUCTION

The most important asset of an organization is its data. Different forms of data can be gathered through multiple sources. Big data refers to a huge amount of structured and unstructured data. This data is used by big organizations as important information. It has paved its way in almost all sectors important decisions are made based on big data. Big data is mostly contained the latest information and has a growing nature [1][2]. Organizations use this data to know the needs, wants, and preferences of their customers. The biggest challenge faced by organizations is to handle big data. Big data domination uses different tools and techniques for managing data. Traditional systems are not capable enough to handle complex data so it requires advanced systems for data handling. A big data governance framework is a set of guidelines for handling large amounts of data. The governance model is designed depending on the structure of an enterprise and its needs. It has tremendous benefits along with some challenges that need to be overcome especially in implementing the big data power model [3].

This research is an illustration of the position of big data governance and its framework. For handling big data in a company, a well-organized big data governance structure is necessary. Big data governance frameworks are used to guide the management of large amounts of data. The tools and techniques used for big data governance are also mentioned. Together with that, the issues and benefits associated with big data are discussed in the report. data governance, its framework, pros and cons, and governance in practice. Section 3 shows a case study of the big data supremacy National pension service of South Korea. In the last section, A detailed conclusion provides future insight into big data governance.

A. BACKGROUND

Data is useful in almost every way in an organization. Technological advancements are increasing day by day. Traditional ways of getting data are not as helpful as they do not tend to provide the data with such accurate maintenance of data.The operations of organizations are separated into different parts that are integrated. The main issues that arose due to the use of traditional data management systems were related to data security and privacy and data storage. These complex problems require smart solutions. Big data comprises of 5Vs that are velocity, volume, variety, veracity, and value. These all are managed through a big data supremacy model by providing high-end security and reliability of data. Cloud computing is another latest way of data storage with high-security concerns.

A. PROBLEM STATEMENT

Importance of big data power

B. RESEARCH OBJECTIVES (RO)

The main aims of conducting the research are:

RO1: To find out the importance of big data and its governance.
RO2: To discuss the big data domination frameworks.
RO3: To highlight the tools required for implementing big data governance.
RO4: To investigate the opportunities and challenges regarding big data governance.

C. RESEARCH QUESTIONS (RQ)

The key research topics have been identified to carry o ut this SLR successfully. Furthermore, a detailed search method for identifying and extracting the most important articles has

been created as part of the assessment. Table I lists the research questions covered in this review, along with their main rationale.

Table 1: RQ and major motivations

	Research Question	*Major Motivation*
RQ1	How does big data use big data governance technologies for carrying out business needs?	To understand how technologies have been used to investigate and transform data.
RQ2	What is the governance structure for big data?	To understand guiding principles that help in the governance of data.
RQ3	What are the advantages and issues associated with big data governance?	To identify advantages and issues with big data governance.
RQ4	Which best practices are used for big data governance?	To identify and understand best practices used for big data governance.

D. RESULTS AND ANALYSIS

The findings of the research paper show that big data governance has been applied in almost all sectors as it provides a significant way of managing the data. There are eight elements in the big data framework and different types of tools for implementing data governance some of them are given. The opportunities regarding big data governance implications are data security, integrity, storage, transparency, etc. and the challenges include cost, maintenance, and leadership, etc.

II. METHODOLOGY

The method used to carry out the research is a secondary method of data collection. There are two basic ways of gathering data. Those are primary data collection and the secondary data collection method. Primary data is first-hand data that can be taken out through experiments and surveys. Secondary data refers to collecting already available data such as through the internet or books etc. there are different types of data such as qualitative data and quantitative data both can be gathered through different data collection techniques. In this report, a qualitative data type is used. To support our analysis we will consider scholarly articles, journals, and online resources. The research questions are answered in the paper.

BIG DATA GOVERNANCE FRAMEWORK, MOTIVATION AND RELATED STUDIES

RQ1: How does big data use big data governance technologies for carrying out business needs?

A Big data refers to large sets of structured and unstructured data used to gather required information by businesses. Big data is very difficult to handle especially in terms of its privacy and storage. Big data refers to enormous data collections with a diverse and complicated structure that are challenging to store, analyses, and visualize for subsequent processes or outcomes [20]. It may also include confidential information of the customers. There are laws made regarding the security of big data that is why it is really important to manage big data properly. In recent years, creating Big Data applications has become increasingly vital. In fact, a growing number of businesses from many industries rely on knowledge extracted from massive amounts of data [21]. Many important decisions are made based on big data especially those that are customer-oriented. Huge Data Analytics has given rise to a new paradigm and set of solutions for big data sources, storage, and sophisticated analytics [22]. Big data is being increasingly worried, created, and employed for all professions and industries due to its large amount and complexity [23]. Rules and regulations developed for managing big data are called the governance of data [4][4] [4].

A. BIG DATA GOVERNANCE AND ITS FRAMEWORK

Big data governance refers to managing big data through following rules and regulations. The Internet has become a prominent and visible problem in global politics due to the importance of digital technology for social and economic development, as well as an increasing focus on data collecting and privacy issues [24]. This will provide a proper way of dealing with big data according to the big data governance framework [16]. Telecommunications operators have a golden opportunity to create new sources of revenue using Big Data Analytics (BDA) solutions with the increase in data traffic due to a change in customer behavior toward the use of telecommunications services, fostered by the current global health situation [25]. Companies, both private and public, have become more efficient as a result of the recent expansion in ICT and digital data [17]. It helps make useful decisions in an enterprise. Many factors are included in the big data framework model. Data governance covers a wide range of topics that require close attention to guarantee that governance is successful [55]. The Big Data governance framework adds to the existing data governance requirements by focusing on data quality. It focuses on providing data services that are timely, reliable, meaningful, and sufficient [18]. To identify gaps in existing frameworks, the selected studies (i.e. BGF1 to BGF12) on Big Data frameworks were compared to the ISO 8000 data governance framework standard. Table 2 compares the ISO 8000 standard, where FR stands for full representation and PR for partial representation. Because the BGF7 framework is not available, no comparison can be made. The rapid growth of regional health information networks (RHINs) around the world is being hampered by the emergence of big data [19]. The practical implication of big data governance includes many challenges and opportunities [5].

RQ2: What is the governance structure for big data?

Big data governance frameworks are the guiding principles that help in the governance of data. It includes instruction criteria for carrying out data governance. By following the data framework, a smooth implication of big data is done in an organization. Figure 1 shows the eight components of a big data governance framework. Big data have a great impact on all operations of an enterprise [6].

Table 2: Compares Big Data governance frameworks to the ISO 8000 standard.

8000 ISO framework governance data	Strategy	Procedure for Authentication	Data on recognition	Identifying Partner	Technology for recognition	Execution and analysis	Quality To the conformance	(%)(FRConclusion)
BGF1	FR	FR	FR	FR	FR	PR	PR	70 %
BGF2	PR	PR	FR	FR	PR	FR	PR	40%
BGF3	PR	PR	FR	PR	PR	PR	PR	10%
BGF4	FR	PR	PR	PR	PR	PR	PR	10%
BGF5	PR	PR	PR	PR	PR	PR	FR	10%
BGF6	PR	FR	FR	PR	PR	FR	FR	67%
BGF7	NA	NA	NA	NA	NA	NA	NA	NA
BGF8	PR	FR	FR	PR	PR	FR	FR	67%
BGF9	FR	FR	PR	PR	PR	PR	PR	27%
BGF10	FR	FR	FR	PR	PR	PR	PR	45%
BGF11	FR	FR	FR	FR	PR	PR	PR	59%
BGF12	PR	FR	FR	FR	PR	PR	PR	43%
Future framework	FR	FR	FR	FR	FR	PR	FR	91%

Different organizations have different types of data frameworks according to their structure, goals, and vision. Big data governance framework includes eight components that are given below:

1) **IDENTIFY THE STRUCTURE OF AN ORGANIZATION**

Before developing the data framework of an organization it is important to know about the structure of an organization based on which data framework is designed. Big data is a term used to describe data that is enormous in size, diverse in diversity, and has a truly tough multiple structure. It is generated through internet applications and communication [60]. It serves differently for different types of organizations [7].

2) **DATA QUALITY AND MEASUREMENT**

For big data adoption to be effective, data quality should be assessed. The era of big data has come in the process industry, due to the ever-increasing amount of data collected from processes [34]. The term 'Big Data' characterizes data by its volume, and also by its velocity, variety, and veracity [35]. This technology has the potential to become a strategic stimulant for businesses seeking a competitive advantage and long-term success [36]. City characteristics and outcome factors may now be measuzred at higher collection rates and at more accurate geographic scales than ever before due to recent "big data" sources [37]. Artificial intelligence techniques give more accurate, faster, and scalable results in big data analytics than traditional data techniques and platforms [38]. The model is created based on the data's quality and worth.

3) **OPTIMIZE AND COMPUTE**

Optimizing and computing means extracting the required information from a large set of data. With rapid growth of smart device users comes an increase in volume of data generated from various smart devices, which varies according to all fundamental V's used to classify it as big data [39]. The goal of optimization is to find the best method given a set of restrictions, such as optimising factors like efficiency, productivity, reliability, longevity, strength, and usage [40]. Optimization of task parallelism of applications in such environments is one of the important components that can handle this challenge [41].

The suggested design uses the infrastructure-as-code paradigm to allow for dynamic cluster configuration and management [42]. Many standards and real-world multi-objective optimization problems have shown that multi-objective evolutionary algorithms (MOEAs) perform well [43]. Not all the data is important so it is important to gather only data that is necessary. The data governance model can help the best for this.

4) **SCOPE OF BIG DATA**

Big data scope helps in finding out the insight into the required data or the definition of the data. It is designed for the organization's needs. In medical applications, data analytics is used to extract important information from large amounts of data samples [44]. The use of Big Data analytics in healthcare has enormous potential for increasing patient care quality, reducing waste and error, and lowering costs [45]. Big Data Analytics was created to identify patterns in massive amounts of data and generate important insights from it [46]. The existing literature implies that customers are primarily passive data providers and that companies are in charge of big data analytics [47]. Although firms now have access to vast volumes of data, simply having the data is not enough; extensive data analysis is required to make better business decisions [48]. Several research have looked into the potential use of BDA in healthcare [49]. Rather than some fundamental issues, big data analytics can provide learners with tailored learning environments, which can reduce possible dropouts, lower academic risks, reduce complexity, and improve the overall quality of the educational system [50]. Utility companies can benefit from Big Data technologies, but deciding which Big Data technology to deploy is crucial [52]. By considering the data scope, the big data governance framework model is designed.

5) **RULES AND POLICIES**

Data governance framework is an overall set of procedures and policies designed for managing big data. It is designed to do things in a designed pattern.

6) **DATA MANAGEMENT**

The main component of the data governance framework is data management. The availability of new computational models in the field of healthcare, together with advancements in medical big data, has enabled caregivers and researchers to extract useful information and visualise healthcare big data in a new spectrum [51]. The information's dependency is classed based on the attributes-based linkages modelled between the data [53]. In biomedical informatics, big data management for information standardization and integration in health research is a defining challenge [54]. For the management of big data proper framework is required and it is not so easy to deal with big data.

7) **STAKEHOLDER'S SELECTION**

Stakeholder selection means deciding to which stakeholders the access to the data must be given. There are many stakeholders linked with an organization. Some stakeholders are close to the organization's internal matters they need to examine the data for making decisions [8].

8) **STORAGE OF BIG DATA**

Data storage is one of the biggest challenges regarding big data. This issue can be solved by using big data governance. A cloud-based big data sharing system will make use of a cloud service provider's storage facility to exchange data with authorised users belowbelowbelow below[26]. Big data overall provides the why, who, how, when, and what of a business. With the broad use of Internet-enabled gadgets, information sharing across geographically dispersed smart devices has increased exponentially[27]. Big data-based acquisition and storage system plays an essential part in the design of industrial data platform [28]. Cloud storage allows programmes to handle their remote data more efficiently, but it comes with the risk of being tampered with [29]. Big data systems are sufficiently stable to store and process a massive volume of rapidly changing data [30]. Checking the integrity of cloud data shared by a group of users is referred to as cloud storage auditing systems for shared data [31]. With the growing use of smart and Internet-connected objects in a variety of applications, there is a commensurate increase in data transfer across different geographic regions [32]. Big data is a collection of structured, semi-structured, and unstructured data on a massive scale [33]. There are many other principles regarding big data framework some of them include Metadata, privacy policy or security, people, quality, etc. big data plays an important role in making decisions and constructing policies in an organization. Big data governance framework act as a road map for guiding the adoption of big data governance techniques by using different tools. Both small and large enterprises use big data for extracting useful information. Big data governance has evolved the traditional ways of executing data. While technology for establishing and operating big data projects have improved and spread in recent years, fully using big data's potential is still in its infancy. In actuality, the term "big data" refers to large data sets with a high velocity, volume, and variety, as well as a complicated structure that makes management, analysis, storage, and processing challenging [64]. The efficiency of storage systems can be greatly enhanced by finding the cost of data. Cognitive storage is a crucial idea, which involves better understanding the importance of data to user needs and preferences in order to optimise storage systems [65].

Large data storage technologies, that were used to construct a bridge roadmap for big data technology development in a variety of high-impact application sectors**Error! Reference source not found.**. Big data has now become the most difficult problem in the industrial, scientific, and educational sectors. In this section, we'll talk about the storage issues that these industries face.

In recent years, data has grown. And how we're managing with our industries' huge amounts of data [67][68][69]-[80].

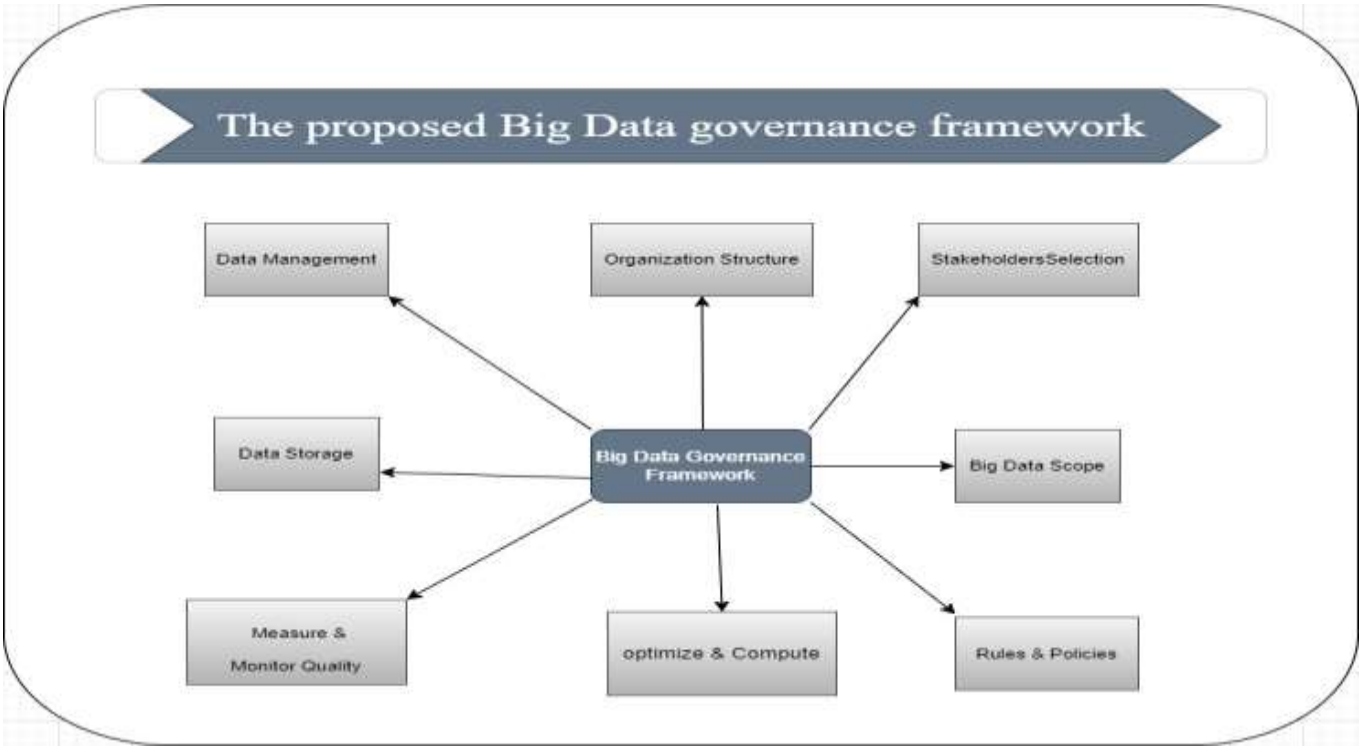

Figure 1: The proposed Big Data governance framework

B. TOOLS FOR BIG DATA GOVERNANCE

APACHE ATLAS

This is the Hadoop supremacy data structure in general. For metadata, Hadoop components are governed via data governance. The notion of tagging is utilized to ensure security [9].

1) ALATION DATA CATALOG

A data catalog that provides access that is single-sourced to multiple sources to gather data for consumers is known as an Alation data catalog. It regulated the activities such as customer engagement, updating, and collaboration used in data governance [10].

2) SAP MASTER DATA GOVERNANCE

SAP master data governance enables a company to assure data quality in combination with rules and regulations used for the management of data. In the meantime, text analysis reveals that 1) master data, 2) data quality, 3) business intelligence, 4) business process, 5) data integration, 6) big data, 7) data governance, 8) information governance, 9) data management, and 10) product data are common terms and interest topics in the Master Data Management research [56]. Metadata aids in the identification of critical data for an organization. Different terminology is also used to make it easier to locate the information needed.

RQ3: What are the advantages and issues associated with big data governance?

III. CHALLENGES AND OPPORTUNITIES IN BIG DATA GOVERNANCE

Big data governance is a necessary procedure to follow, but it comes with a slew of possibilities and problems that a firm must be aware of before implementing it. Figure 2 demonstrates some of the possibilities in big data governance. Data governance provides benefits such as data security, cost savings, and so on.

A. OPPORTUNITIES

Some of the opportunities in big data governance are mentioned below:

1) DATA TRANSPARENCY

Data governance ensures data transparency. Software can handle more complicated jobs, such as identifying fraud, optimising logistics routes, and even driving cars, thanks to big data and clever algorithms [57]. It is considered that, for the time being, full openness for oversight bodies is the only viable choice; extending it to the general public is usually not recommended [58]. When data governance policies are implemented in a company, the visibility of the organization's activities improves. Data transparency aids in recognizing and finding important data from relevant websites. A significant quantity of data, particularly about consumers, is accumulated in business databases, which are extremely difficult to handle. As a result, a huge quantity of data is generated, including sensitive information that may be in danger of being shared due to rising data theft. Organizations like to secure data by utilizing various software programs. Data governance aids in data security by establishing privacy policies and security

standards to prevent data loss. Database systems are more safeguarded as a result of the increased openness of meta-data having controlled access over the data [11].

2) IMPROVED DATA SECURITY AND PRIVACY

These days, data security is regarded as the most essential factor, and businesses strive to improve data security daily to maintain consumer confidence. It is a complex process of securing data but with the emergence of big data governance, data security and privacy are under control as more laws and regulations to protect the data are being enacted. Cloud computing allows for easy data transfer while maintaining security and allowing for thorough analysis. Data is accessible in a secure manner. It is extremely easy to utilize inside the company to provide a safer platform. This is a fantastic chance to securely gather and store massive amounts of data in databases using big data governance.

3) CUSTOMERS CENTRIC

Customer-focused businesses gather as much information on their clients as possible from various sources. It means that businesses are getting more customer-centric by the day. They are putting forth a lot of effort to provide more interactive interfaces, such as websites. An excellent shift is to teach people how to use machinery.Customers and businesses benefit from data governance because it makes it simpler to locate what they want in less time. Customer happiness and revenues grow as a result of offering a valuable experience. Businesses must have access to clients and their needs if they are expanding.

4) LIMITED ACCESS OF DATA

The adoption of data governance ensures secure access to the customer's data, which has become one of the most pressing problems. When a significant amount of data is transferred from one location to another, there is a risk of data leakage. Customers are granted access to customer portals that are only available to approved customers. Data access is governed by security laws. To these security issues, data governance provides a solution and improved techniques.The following are some of the difficulties connected with data governance, along with their solutions:

5) HANDLING OF DATA

Different organizations, such as CCAR, have implemented various regulatory procedures to handle data. In all departments, these regulations should be uniform and regulated properly. Organizations must develop structures according to definite rules. It proves the governance in an organization. The main challenge is to make these rules according to the laws imposed within a region [12].

6) IRREGULARITY IN DATA INTEGRITY

Data integrity means the correctness of data with consistency that is important for businesses, especially when working with large amounts of data. Organizations must employ advanced data collection strategies. Organizations are all battling for data integrity. The most difficult aspect of data governance is ensuring data integrity for various types of businesses. It creates new data accuracy policies and laws. As a result, the data becomes a valuable asset to the company. It develops consumer trust. It is a big challenge to ensure big data integrity.

7) DATA CENTRALIZATION

Data centralization is difficult for companies since it requires storing data in a data hub or Data Lake, which necessitates the adoption of data governance. Big data has become the driving force behind the evolution of many machine learning advances in recent years as data has spread to stratospheric levels [59]. It takes time and effort to examine data and its source or nature, before moving it. All efforts would be wasted if data governance is not managed properly. As a result of the discussion, data governance is vital for providing data value.

8) IDENTIFY STAKEHOLDER'S ROLE

When a data governance model is created, identifying the roles of stakeholders and workers may be a challenging process for a company. Most of the time employees do not welcome change in the technology which inhibits communication between stakeholders which causes problems. Before any incident occurs, roles and responsibilities must be established in line with the governance framework to execute an effective governance model. To avoid wasting time, future tasks must be given on time. The chief data officer is in charge of the governance model roles and responsibilities together with other management issues [13].

RQ4: Which best practices are used for big data governance?

IV. DATA GOVERNANCE IN PRACTICE

Many businesses, particularly customer-oriented organizations, employ data governance. For the firm's success, implementing data governance is a large commitment that will take time and effort. In reality, four crucial stages are followed. After that, a solid structure with policies and procedures is established by forming a team. Big businesses must deal with a large volume of data. Data governance is also used to aid with innovation and technology management. If properly handled, data may become an organization's most important asset, assisting it in remaining competitive and flexible, predicting customer demands, and controlling costs [61]. On their own, data management systems are becoming prohibitively expensive and unable of facing the reality of ever-increasing data complexity. Businesses have become more complex in their use of data, resulting in new demands that require new approach in data management. The only way to solve the data problem, according to forward-thinking corporations, is to develop an effective data governance system [62]. Many firms have realized that customer feedback is a significant corporate asset that should be carefully protected and its value actively managed or 'regulated.' An increasing number of individuals are learning the lesson that data can also be a major liability [63].The best practices for big data governance are given below:

- The data governance must be simple so that every individual can understand it.
- Use different matrices and report for measuring the success of the data

- It is important to shift towards automation as much it is possible to avoid errors [14]
- Data governance teams must be assigned different roles for creating and managing the data

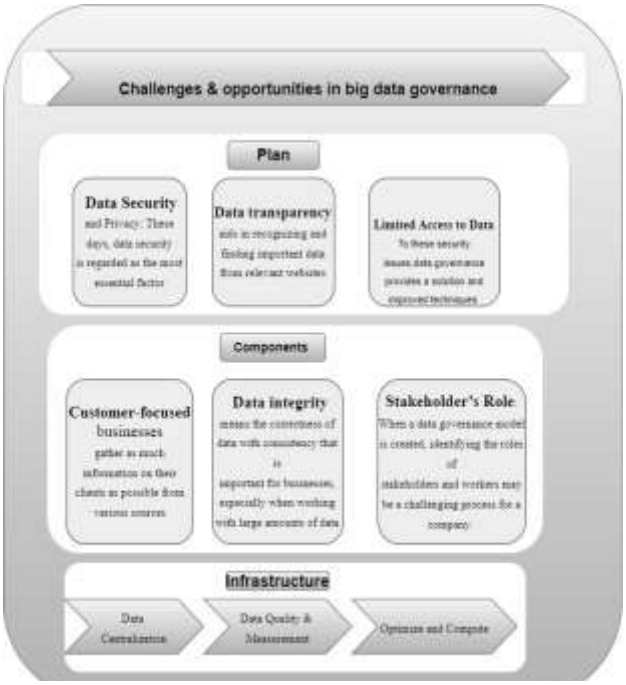

Figure 2 :Challenges and opportunities in big data governance

V. CASE STUDY PRESENTING BIG DATA GOVERNANCE NATIONAL PENSION SERVICE OF SOUTH KOREA

South Korea's national pension service is one of the world's largest pension institutions. It has a large number of assets and a large channel. It is now growing into an open data policy with the government of Korea, which is known as government 3.0. Its public data promotion effort is this trend toward public data. The information released is big data, which is critical to collect since it will open doors for new enterprises to grow. In 2013, over 20 terabytes of data were stored. The organization's growth needs to understand public trends hence, new approaches, such as big data analysis, are necessary to collect such data to improve customer value [15]. Unstructured and structured data is gathered from several sources for this aim. Consequently, the public's confidence was earned, and an overwhelming reaction was witnessed, resulting in the system's efficiency being attained. The major goal is to administer huge data efficiently and provide the best service possible. Two departments have been created to ensure the security of big data. The information protection and personal information protection departments are in charge of this. These are important in establishing rules and legislation for data protection policies that prevent a data breach. The accuracy and quality of data are also preserved while data security is increased. In addition, timeliness is noted concerning the big data governance structure. With big data governance, managing such a large network has become a simple process. The data's reliability, regularity, and importance are all positively impacted NPS's systems have had a great deal of success in their development. The organization's meta-data is structured consistently and systematically. In 2016, NPS created a full-scale big data governance strategy in response to its fast success [15].

VI. CONCLUSION

Finally, the necessity of adopting big data governance to enhance the efficiency of corporate processes is discussed. Data governance offers rules and standards for data organization since managing a vast volume of data has become a significant issue in the corporate sector. Data governance, which is at the base of data management in companies, ensures data consistency, dependability, and correctness. Domains, which are essentially the model of data governance used in businesses, are one of the many elements of big data. Data governance has become a necessity of today's era because it provides numerous benefits regarding data management. Big data governance has an enormous amount of opportunities like data transparency, security, and CRM. There are also some challenges like data centralization, data integrity, and regulatory compliance, among others. The application of the finest data governance model is the answer to all problems. Different technologies and techniques for data governance are used for efficient deployment, it is highly essential to understand. There is a well-established framework for efficient communication between the firm and its partners. Organizations must always encourage technological progress and innovation to keep these high-tech systems running smoothly, and all needs must be fulfilled on time to maintain data consistency. The process of adopting data governance in a company is extensive, and it takes time and effort to integrate it into the workplace. Managing staff roles and responsibilities is also a crucial responsibility. Because traditional systems are incapable of handling such vast amounts of data, the new data governance model must become ingrained in the corporate culture for creative breakthroughs to occur.

CREDIT AUTHOR STATEMENT: Mansoor Ahmed Rasheed: Conceptualization, Methodology, draft handling and maintain, Abstract, Introduction, Big data governance and its Framework, Figures. **Hafiz Zeeshan Ahmad**: Introduction, Background, Motivation and Related Studies, Research questions, **Hafiz Abdullah Tanweer**: Results and analysis, Research questions,Methodology, Irregularity in Data Integrity, Case study presenting big data governance.**Hudabia Murtaza:**Methodology,Table making and Editing, Optimize and Compute, Rules and Policies, Data Management, Scope of big data. **Mannan Ahmad Rasheed**: Stakeholder's Selection, Storage of Big Data, Tools for big data Governance, Mishaal **Ahmed:**Opportunities, Improved Data Security and Privacy, Data centralization.

COMPLIANCE WITH ETHICAL STANDARDS: It is declare that all authors don't have any conflict of interest.Furthermore, informed consent was obtained from all individual participants included in the study

REFERENCES

[1] S. Sagiroglu, D. Sinanc, "Big data: A review," in 2013 International Conference on Collaboration Technologies and Systems (CTS), 2013.

[2] V. Morabito, "Big data governance," in Big data and analytics, 2015, pp. 83-104.

[3] G. George, Haas, M. R., & Pentland, " Big data and management," Academy of Management JournalVol. 57, No. 2, pp. 64-72, 2014.

[4] S. Soares, Big data governance. Information Asset, LLC. 2012.

[5] A. T. Ali Al-Badi, Asharul Islam Khan, "Exploring big data governance frameworks," Procedia Computer Science, pp. 271-277, 2018.

[6] P. Ghavami, "The Key Components of Big Data Governance," in Big Data Management, 2020, pp. 64-72.

[7] M. P. Atsushi Yamada, "Governance framework for enterprise analytics and data," in 2017 IEEE International Conference on Big Data, 2017.

[8] M. Tang, Shao, Saisai, Yang, Weiqing, Liang, Yanbo, Yu, Yongyang, Saha, Bikas, Hyun, Dongjoon, "A System for Big Data Lineage Tracking," in IEEE 35th International Conference on Data Engineering (ICDE), Macao, 2019.

[9] D. Well, Introduction to Data Catalogs. Alation, 2019.

[10] S. D. Neil D Lawrence, Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. International Data Privacy Law, Volume 9, Issue 4, November 2019, Pages 236–252, 2019.

[11] J. S. P. Olivia Benfeldt, Sabine Madsen, "Data Governance as a Collective Action Problem," Information Systems Frontiers, pp. 299-313, 2020.

[12] C.-k. C. Daniel W. K. Tse, "The Challenges of Big Data Governance in Healthcare," in 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2018.

[13] Z. Panian, "Some Practical Experiences in Data Governance," in World Academy of Science, Engineering and Technology, 2010.

[14] H. Y. Kim and J.-S. Cho, "Data governance framework for big data implementation with NPS Case Analysis in Korea," Journal of Business and Retail Management Research, vol. 12, no. 3, 2018.

[15] S. Sagiroglu and D. Sinanc, "Big data: A review," in 2013 international conference on collaboration technologies and systems (CTS), 2013: IEEE, pp. 42-47.

[16] H. Y. Kim and J.-S. Cho, "Data governance framework for big data implementation with a case of Korea," in *2017 IEEE International Congress on Big Data (BigData Congress)*, 2017: IEEE, pp. 384-391.

[17] A. Al-Badi, A. Tarhini, and A. I. Khan, "Exploring big data governance frameworks," *Procedia computer science*, vol. 141, pp. 271-277, 2018.

[18] H. Y. Kim and J.-S. Cho, "Data governance framework for big data implementation with NPS Case Analysis in Korea," *Journal of Business and Retail Management Research*, vol. 12, no. 3, 2018.

[19] Q. Li *et al.*, "A framework for big data governance to advance RHINs: a case study of China," *IEEE Access*, vol. 7, pp. 50330-50338, 2019.

[20] S. Sagiroglu and D. Sinanc, "Big data: A review," in *2013 international conference on collaboration technologies and systems (CTS)*, 2013: IEEE, pp. 42-47.

[21] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big Data technologies: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 431-448, 2018.

[22] M. K. Saggi and S. Jain, "A survey towards an integration of big data analytics to big insights for value-creation," *Information Processing & Management*, vol. 54, no. 5, pp. 758-790, 2018.

[23] M. Ge, H. Bangui, and B. Buhnova, "Big data for internet of things: a survey," *Future generation computer systems*, vol. 87, pp. 601-614, 2018.

[24] M. Flyverbom, R. Deibert, and D. Matten, "The governance of digital technology, big data, and the internet: New roles and responsibilities for business," *Business & Society*, vol. 58, no. 1, pp. 3-19, 2019.

[25] M. Z. Kastouni and A. A. Lahcen, "Big data analytics in telecommunications: Governance, architecture and use cases," *Journal of King Saud University-Computer and Information Sciences*, 2020.

[26] S. Han, K. Han, and S. Zhang, "A data sharing protocol to minimize security and privacy risks of cloud storage in big data era," *IEEE Access*, vol. 7, pp. 60290-60298, 2019.

[27] G. S. Aujla, R. Chaudhary, N. Kumar, A. K. Das, and J. J. Rodrigues, "SecSVA: secure storage, verification, and auditing of big data in the cloud environment," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 78-85, 2018.

[28] D. Geng, C. Zhang, C. Xia, X. Xia, Q. Liu, and X. Fu, "Big data-based improved data acquisition and storage system for designing industrial data platform," *IEEE Access*, vol. 7, pp. 44574-44582, 2019.

[29] J. Li, J. Wu, G. Jiang, and T. Srikanthan, "Blockchain-based public auditing for big data in cloud storage," *Information Processing & Management*, vol. 57, no. 6, p. 102382, 2020.

[30] M. Saadoon, S. H. A. Hamid, H. Sofian, H. H. Altarturi, Z. H. Azizul, and N. Nasuha, "Fault tolerance in big data storage and processing systems: A review on challenges and solutions," *Ain Shams Engineering Journal*, 2021.

[31] Y. Zhang, J. Yu, R. Hao, C. Wang, and K. Ren, "Enabling efficient user revocation in identity-based cloud storage auditing for shared big data," *IEEE Transactions on Dependable and Secure computing*, vol. 17, no. 3, pp. 608-619, 2018.

[32] A. Singh, S. Garg, K. Kaur, S. Batra, N. Kumar, and K.-K. R. Choo, "Fuzzy-folded bloom filter-as-a-service for big data storage in the cloud," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2338-2348, 2018.

[33] G. Bathla, R. Rani, and H. Aggarwal, "Comparative study of NoSQL databases for big data storage," *International Journal of Engineering & Technology*, vol. 7, no. 2.6, pp. 83-87, 2018.

[34] L. Yao and Z. Ge, "Big data quality prediction in the process industry: A distributed parallel modeling framework," *Journal of Process Control*, vol. 68, pp. 1-13, 2018.

[35] S. Juddoo, C. George, P. Duquenoy, and D. Windridge, "Data governance in the health industry: Investigating data quality dimensions within a big data context," *Applied System Innovation*, vol. 1, no. 4, p. 43, 2018.

[36] M. Wook *et al.*, "Exploring big data traits and data quality dimensions for big data analytics application using partial

least squares structural equation modelling," *Journal of Big Data*, vol. 8, no. 1, pp. 1-15, 2021.
[37] E. L. Glaeser, S. D. Kominers, M. Luca, and N. Naik, "Big data and big cities: The promises and limitations of improved measures of urban life," *Economic Inquiry*, vol. 56, no. 1, pp. 114-137, 2018.
[38] R. H. Hariri, E. M. Fredericks, and K. M. Bowers, "Uncertainty in big data analytics: survey, opportunities, and challenges," *Journal of Big Data*, vol. 6, no. 1, pp. 1-16, 2019.
[39] R. Chaudhary, G. S. Aujla, N. Kumar, and J. J. Rodrigues, "Optimized big data management across multi-cloud data centers: Software-defined-network-based analysis," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 118-126, 2018.
[40] N. Das, L. Das, S. S. Rautaray, and M. Pandey, "Big data analytics for medical applications," *International Journal of Modern Education and Computer Science*, vol. 11, no. 2, p. 35, 2018.
[41] Á. B. Hernández, M. S. Perez, S. Gupta, and V. Muntés-Mulero, "Using machine learning to optimize parallelism in big data applications," *Future Generation Computer Systems*, vol. 86, pp. 1076-1092, 2018.
[42] M. Grzegorowski, E. Zdravevski, A. Janusz, P. Lameski, C. Apanowicz, and D. Ślęzak, "Cost optimization for big data workloads based on dynamic scheduling and cluster-size tuning," *Big Data Research*, vol. 25, p. 100203, 2021.
[43] H. Wang *et al.*, "A hybrid multi-objective firefly algorithm for big data optimization," *Applied Soft Computing*, vol. 69, pp. 806-815, 2018.
[44] S. Mishra, B. K. Mishra, H. K. Tripathy, and A. Dutta, "Analysis of the role and scope of big data analytics with IoT in health care domain," in *Handbook of data science approaches for biomedical engineering*: Elsevier, 2020, pp. 1-23.
[45] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," *International journal of medical informatics*, vol. 114, pp. 57-65, 2018.
[46] N. J. Ogbuke, Y. Y. Yusuf, K. Dharma, and B. A. Mercangoz, "Big data supply chain analytics: ethical, privacy and security challenges posed to business, industries and society," *Production Planning & Control*, pp. 1-15, 2020.
[47] H. Zhang and Y. Xiao, "Customer involvement in big data analytics and its impact on B2B innovation," *Industrial Marketing Management*, vol. 86, pp. 99-108, 2020.
[48] J. R. Saura, B. R. Herráez, and A. Reyes-Menendez, "Comparing a traditional approach for financial Brand Communication Analysis with a Big Data Analytics technique," *IEEE Access*, vol. 7, pp. 37100-37108, 2019.
[49] S. Khanra, A. Dhir, A. N. Islam, and M. Mäntymäki, "Big data analytics in healthcare: a systematic literature review," *Enterprise Information Systems*, vol. 14, no. 7, pp. 878-912, 2020.
[50] P. M. Titimus, "Big Data Analytics in the Higher Education: Need of the Future," in *Advances in Communication, Cloud, and Big Data*: Springer, 2019, pp. 23-28.
[51] S. Nazir *et al.*, "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95714-95733, 2020.
[52] H. Daki, A. El Hannani, A. Aqqal, A. Haidine, and A. Dahbi, "Big Data management in smart grid: concepts, requirements and implementation," *Journal of Big Data*, vol. 4, no. 1, pp. 1-19, 2017.
[53] G. Manogaran *et al.*, "FDM: Fuzzy-optimized data management technique for improving big data analytics," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 177-185, 2020.
[54] X. Wang, C. Williams, Z. H. Liu, and J. Croghan, "Big data management challenges in health research—a literature review," *Briefings in bioinformatics*, vol. 20, no. 1, pp. 156-167, 2019.
[55] J. M. Barker, *Data Governance: the missing approach to improving data quality*. University of Phoenix, 2016.
[56] F. Haneem, R. Ali, N. Kama, and S. Basri, "Descriptive analysis and text analysis in systematic literature review: a review of master data management," in *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, 2017: IEEE, pp. 1-6.
[57] H. Schildt, "Big data and organizational design–the brave new world of algorithmic management and computer augmented transparency," *Innovation*, vol. 19, no. 1, pp. 23-30, 2017.
[58] P. B. De Laat, "Algorithmic decision-making based on machine learning from Big Data: Can transparency restore accountability?," *Philosophy & technology*, vol. 31, no. 4, pp. 525-541, 2018.
[59] H. Matallah, G. Belalem, and K. Bouamrane, "Towards a new model of storage and access to data in big data and cloud computing," *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 8, no. 4, pp. 31-44, 2017.
[60] M. A. Praveena and B. Bharathi, "A survey paper on big data analytics," in *2017 International Conference on Information Communication and Embedded Systems (ICICES)*, 2017: IEEE, pp. 1-9.
[61] Z. Panian, "Some practical experiences in data governance," *World Academy of Science, Engineering and Technology*, vol. 62, no. 1, pp. 939-946, 2010.
[62] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, "A systematic literature review of data governance and cloud data governance," *Personal and Ubiquitous Computing*, vol. 23, no. 5, pp. 839-859, 2019.
[63] A. Gregory, "Data governance—Protecting and unleashing the value of your customer data assets," *Journal of Direct, Data and Digital Marketing Practice*, vol. 12, no. 3, pp. 230-248, 2011.
[64] K. A. I. Hammad, M. A. I. Fakharaldien, J. M. Zain, and M. Majid, "Big data analysis and storage," in *International Conference on Operations Excellence and Service Engineering*, 2015, pp. 10-11.
[65] K. Krish, B. Wadhwa, M. S. Iqbal, M. M. Rafique, and A. R. Butt, "On efficient hierarchical storage for big data processing," in *2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, 2016: IEEE, pp. 403-408.
[66] G. Cherubini, J. Jelitto, and V. Venkatesan, "Cognitive storage for big data," *Computer*, vol. 49, no. 4, pp. 43-51, 2016
[67] M. Strohbach, J. Daubert, H. Ravkin, and M. Lischka, "Big data storage," in *New horizons for a data-driven economy*: Springer, Cham, 2016, pp. 119-141.
[68] M. Padgavankar and S. Gupta, "Big data storage and challenges," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, pp. 2218-2223, 2014.
[69] A. Arooj, M. S. Farooq, A. Akram, R. Iqbal, A. Sharma, and G. Dhiman, "Big Data Processing and Analysis in Internet of

Vehicles: Architecture, Taxonomy, and Open Research Challenges," *Archives of Computational Methods in Engineering*, pp. 1-37, 2021.

[70] A. H. A. Zahid, M. W. Haider, M. S. Farooq, A. Abid, and A. Ali, "A critical analysis of software failure causes from project management perspectives," *VFAST Transactions on Software Engineering*, vol. 6, no. 1, pp. 62-68, 2018.

[71] M. S. Farooq, M. Khan, and A. Abid, "A framework to make charity collection transparent and auditable using blockchain technology," *Computers & Electrical Engineering*, vol. 83, p. 106588, 2020.

[72] A. Arooj, M. S. Farooq, A. Akram, R. Iqbal, A. Sharma, and G. Dhiman, "Big Data Processing and Analysis in Internet of Vehicles: Architecture, Taxonomy, and Open Research Challenges," Archives of Computational Methods in Engineering, pp. 1-37, 2021.

[73] A. H. A. Zahid, M. W. Haider, M. S. Farooq, A. Abid, and A. Ali, "A critical analysis of software failure causes from project management perspectives," *VFAST Transactions on Software Engineering*, vol. 6, no. 1, pp. 62-68, 2018.

[74] A. Abid, M. F. Manzoor, M. S. Farooq, U. Farooq, and M. Hussain, "Challenges and Issues of Resource Allocation Techniques in Cloud Computing," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 7, pp. 2815-2839, 2020.

[75] I. A. Khawaja, A. Abid, M. S. Farooq, A. Shahzada, U. Farooq, and K. Abid, "Ad-Hoc Collaboration Space for Distributed Cross Device Mobile Application Development," IEEE Access, vol. 8, pp. 62800-62814, 2020.

[76] A. Arooj, M. S. Farooq, A. Akram, R. Iqbal, A. Sharma, and G. Dhiman, "Big Data Processing and Analysis in Internet of Vehicles: Architecture, Taxonomy, and Open Research Challenges," Archives of Computational Methods in Engineering, pp. 1-37, 2021.

[77] O. Aziz, M. S. Farooq, A. Abid, R. Saher, and N. Aslam, "Research trends in enterprise service bus (ESB) applications: A systematic mapping study," IEEE Access, vol. 8, pp. 31180-31197, 2020.

[78] I. Obaid, M. S. Farooq, and A. Abid, "Gamification for recruitment and job training: model, taxonomy, and challenges," IEEE Access, vol. 8, pp. 65164-65178, 2020.

[79] H. Malik, M. S. Farooq, A. Khelifi, A. Abid, J. N. Qureshi, and M. Hussain, "A Comparison of Transfer Learning Performance Versus Health Experts in Disease Diagnosis From Medical Imaging," *IEEE Access*, vol. 8, pp. 139367-139386, 2020.

[80] R. Tehseen, M. S. Farooq, and A. Abid, "A framework for the prediction of earthquake using federated learning," *PeerJ Computer Science*, vol. 7, p. e540, 2021.

[81] A. Abid, M. S. Farooq, I. Raza, U. Farooq, and K. Abid, "Variants of Teaching First Course in Database Systems," *Bulletin of Education and Research*, vol. 37, no. 2, pp. 9-25, 2015.