IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Quality Assurance in The Age of Data Analytics: Innovations and Challenges

Kishan Patel
Independent Researcher

***Abstract*—**A method for guaranteeing the high standard of software as a whole is quality assurance. Software products utilised for data analytics are distinct from more conventional ones, requiring a particular set of skills and more time for quality assurance. The development and implementation of several big data application systems across different fields have been prompted by the rapid expansion of the big data technology industry, which is now seeing a compound annual growth rate of 27%. The 4Vs of big data are variety, volume, velocity, and veracity, and these systems use a wide array of platforms, tools, and procedures to gather, process, analyse, and forecast this data. However, ensuring the quality of these systems poses significant challenges due to a complexity and scale of a data involved. This paper offers a thorough overview of quality assurance (QA) concepts and components for big data systems, focussing on important criteria including timeliness, consistency, credibility, and accessibility. Validation techniques such as classification-based testing, program-based testing, and metamorphic testing are presented, and the article goes on to talk about the special QA problems caused by the varied and ever-changing nature of big data. Through a systematic literature review, this study synthesises existing QA technologies and their strengths and limitations, offering insights into the current state of QA in big data applications and identifying areas for future research and improvement.

Keywords—Quality Assurance (QA); data analytical, Data Quality Criteria, applications, Methods.

I. INTRODUCTION

Market research indicates that big data technologies have a 27% CAGR, which might lead to over $203 billion in revenue by 2020 [1]. Software systems that can gather, process, analyse, or forecast massive amounts of data using various platforms, tools, and procedures are known as big data application systems [2]. The number of industries making use of big data is growing rapidly, with applications in fields as diverse as statistics, monitoring, and recommendation systems [3][4]. There are so-called "4V" attributes associated with big data applications: variety, volume, velocity, and veracity. Due to the high volume, fast pace of receiving data, and diversity of data kinds, the data generated by big data applications is of poor quality, which in turn compromises the software quality of these applications [5] [6]. Problems with usability and scalability may arise as an outcomes of a proliferation of data kinds and formats, which are becoming more diverse as a result of the vast diversity of heterogeneous data. These formats and types include semi-structured, structured, and unstructured data.

An overarching goal of quality assurance (QA) is to identify and eliminate errors and flaws in software and other produced goods, as well as in the solutions and services that are provided to consumers. In contrast to traditional software systems, four characteristics of big data—the quantity, velocity, correctness, and quality of the data—present substantial challenges for QA technologies when applied to big data [7]. Numerous researchers have highlighted the present QA issues with big data applications [8] [9]. Because of the sheer volume of data and the importance of the data's timeliness, evaluating the availability, accuracy, and performance of a big data forecasting system may be challenging. Constantly scalability in big data recommendation systems is challenging because of the diversity and number of characteristics. As a result, research into quality assurance methods for big data applications is rising in popularity.

A following unique features set big data applications apart from traditional ones: (a) statistical analysis using a wide range of formats and a big amount of organised and unstructured data; (b) intelligent decision-making in the face of ambiguity; (c) knowledge-based system evolution and ML; and (d) demands for more intricate visualisation [10]. To confirm the quality of these new aspects of big data applications, innovative QA tools are required. For instance, big data apps include a great deal of unstructured data in contrast to data in conventional applications (pictures, audio, papers, etc.). The data that is often collected is often diverse and not well-integrated. Traditional testing procedures, particularly performance testing, do not include testing methodologies for large-scale data since such data processing is not necessary in conventional applications [11].

The significance of this paper lies in its exploration of advanced quality assurance (QA) techniques tailored for the complex landscape of big data applications, addressing the unique challenges posed by a vast volume, variety, and velocity of data. Motivated by an increasing reliance on data-driven decision-making across various sectors, the paper aims to highlight the necessity of innovative QA methods to ensure the reliability, accuracy, and integrity of bigdata systems. By examining cutting-edge QA approaches and identifying key data quality elements, the paper seeks to provide a comprehensive framework that can be adopted to enhance the overall effectiveness and trustworthiness of big data applications in critical areas like healthcare, finance, and intelligent systems. This paper's key contributions are:

- Apart from traditional methods, crowdsourcing, learning-based testing, and metamorphic testing are emerging ways to quality assurance for big data systems.
- Ensuring data quality is difficult due to a volume, variety, and velocity of bigdata, rapid data changes, and lack of unified standards.
- Important criteria include completeness, accessibility, integrity, timeliness, readability, authorisation, consistency, credibility, metadata, accuracy, auditability, fitness, and structure.
- Reliable and successful big data applications impacting industries like banking, healthcare, and decision-making need high-quality data.
- To guarantee the performance, dependability, and security of a system, testing methodologies such as program-based, classification-based, data model-based, and rule-based testing are used.

A. Structure of the paper

The structure of the paper begins with an Overview of Quality Assurance, which outlines the fundamental principles and practices essential for maintaining high standards in Section II. Section III then moves on to Data Quality Criteria, which defines high-quality data in the big data environment by outlining its essential components. Section IV, "Challenges of data quality," discusses the problems of guaranteeing data quality as a result of big data's velocity, diversity, and volume. Understanding QA for Big Data Application Systems then explores the scope and processes necessary for QA in these complex systems in Section V. Validation Methods for Big Data Applications discusses various approaches and techniques used to test and validate big data systems in Section VI. The paper also includes a Literature Review to provide context and background on existing research and methodologies in Section VII. Lastly, Section VIII's Conclusion and Future Work section provides a summary of the results and recommendations for further study in the area.

II. OVERVIEW OF QUALITY ASSURANCE

QA is a method of managing the process to ensure that a final product, service, or outcome is up to par in terms of quality and suitable for its intended purpose. Data must be accurate, reliable, and adequate for the intended purpose of any monitoring program or evaluation. The purpose of establishing data quality goals is to provide quantitative and qualitative criteria for building a system that can maintain an acceptable degree of uncertainty within predetermined bounds. The trust of the public and financial bodies will be boosted by quality assurance.

A. Principles of Quality Assurance

Quality Assurance is based on four main concepts.

- **Client Focus:** The goals and requirements of the community and its customers should inform the development of services.
- **Understanding Processes and Systems:** To enhance their services, providers need to understand the service system and its essential service operations.
- **Teamwork:** When problems are identified, solved, and quality is improved as a team, progress is made.
- **Testing Changes the Use of Data:** To find out whether modifications lead to the expected benefit, testing is essential. Processes, challenges, and the efficacy of the adjustments may be better understood with the use of data.

Quality assurance programs in healthcare industries should involve the organizations' medical director, nursing superintendent, paramedical personnel, quality manager and all other dedicated departments. These programs are created to assess healthcare needs, find strategies to make sure that hospital procedures are followed all around a company, have a mechanism to check whether protocols are followed, and launch performance improvement plans when standards aren't reached. Programs for quality assurance often use a quantitative approach to data collection and analysis, which helps pinpoint strengths and weaknesses[12].

III. DATA QUALITY CRITERIA

Due to the novelty of the idea, academic institutions have not yet agreed upon a common definition of "big data" or quality criteria for its use. Data quality is defined differently in different works, but one thing is certain: it is affected by both the data itself and the business environment in which it is used, including the business processes and users. To be deemed qualified, or of high quality, data must adhere to all applicable purposes and specifications. It is common practice for data producers to be the ones who first voice concerns about data quality standards. Data quality was guaranteed in the past since customers were also data providers, either directly or indirectly. But in this big data era, when data comes from all kinds of places, data consumers aren't often data creators.

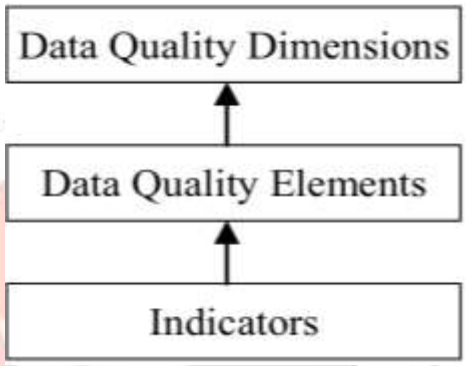

Figure 1: Data quality framework

Measuring data quality is therefore a challenging task. So, as shown in Fig 1, we provide a user-centric hierarchical data quality standard.

A. Data quality elements

A high-quality presentation ensures that readers can comprehend the material to its fullest extent by using accurate terminology to describe it. There are two aspects to it: readability and structure. Data quality components are described below.

1) Accessibility

A data set's accessibility is defined by how difficult it is for consumers to get the data. There is a direct correlation between data openness and accessibility; higher levels of data openness lead to more data kinds collected and more accessibility.

2) Timeliness

The time lag between the creation and collection of data and its use is known as timeliness [13]. This amount of time should be sufficient for data availability, allowing for useful interpretation. Relevance and timeliness are important in an era of big data due to a rapid evolution of data content.

3) Authorization

The term "authorisation" describes the level of permission that a person or group has to use the data.

4) Credibility

Evaluation of non-numerical data is done using credibility. There are two parts to the question of whether a source or message can be believed: the objective and the subjective. The credibility of data is supported by three things: the regularisation of data, a timeliness of data generation, and a reliability of data sources.

5) Definition/Documentation

Data specifications are a part of definitions and documents. They contain things like data names, definitions, valid value ranges, standard formats, business rules, and more. Increased data utilisation is a result of normative data specification.

6) Meta Data

Data usage carries dangers due to data users' distortion of common vocabulary and data concepts, which increases the number of data sources and kinds. To lessen issues brought on by misunderstandings or discrepancies, data providers must include metadata outlining various dataset characteristics.

7) Accuracy

To check whether a data value is accurate, it is compared to a reference value that is already known. For example, when evaluating gender, there are two clear values (male and female) that make it straightforward to gauge accuracy. But there are cases when an acknowledged reference value is lacking, making it hard to assess accuracy.

8) Consistency

The term "data consistency" describes the degree to which associated data follows a proper and comprehensive logic. This often indicates that two separate locations containing the same data should be seen as equal in the context of databases. Data synchronisation refers to the act of achieving parity in data.

9) Integrity

Data integrity is a vague concept that may signify many things depending on the surrounding circumstances. A data type and/or data model dictates how data values are standardised. Dates, definitions, business rules, relations, and any other data attributes must be accurate. Ensuring that data remains accurate and consistent over its full lifecycle is the goal of data integrity in information security.

10)Completeness

A thorough description of quality is possible when a datum includes more than one component. When a single datum is complete, it signifies that all of its parts have valid values.

11)Auditability

There are three stages to the data life cycle, as seen through the lens of audit application: creation, collection, and usage. In this case, nevertheless, auditability implies that auditors may, throughout the data usage phase, properly assess the correctness and integrity of the data within reasonable time and personnel constraints.

12)Fitness

There are two components necessary for fitness: first, the amount of data that users have access to, and second, how well the data created satisfies their needs in terms of indicator definition, elements, categorisation, etc.

13)Readability

Data material is considered readable if it can be accurately described using already established or commonly used words, properties, codes, units, acronyms, or other pieces of information.

14)Structure

Structure, in the context of the vast majority of data being unstructured, is the degree to which it is challenging to use technology to convert semi-structured or unstructured data into structured data.

IV. CHALLENGES of DATA QUALITY

When businesses use and manage big data, the extraction of real, high-quality data from massive, variable, and complicated datasets becomes a critical problem because of the 4V characteristics of big data. The following obstacles are now affecting big data quality:

A. The diversity of data sources brings abundant data types and complex data structures and increases the difficulty of data integration.

Sales and inventory counts, among other data, used to exist exclusively in company systems. The data that companies gather and evaluate nowadays does not adhere to this standard, unfortunately. Among the many diverse sources of big data are: 1) datasets derived by the web and mobile internet; 2) information gathered by the IoT; 3) information gathered by different sectors of the economy; 4) information obtained via scientific investigations and findings, including data from fields such as biology, space travel, and high-energy physics. Rich data types are generated by these sources. Documents, videos, songs, and other forms of unstructured media are examples of data. The second category consists of data that is not strictly organised, such as spreadsheets, financial reports, and software modules. Organised data makes up the third. Eighty percent or more of all data is in the form of unstructured data[14].

The acquisition and successful integration of massive data sets with complex structures from many sources is a formidable challenge for organisations [15]. Data from diverse sources often disagree or have inconsistencies with one another. When dealing with relatively modest amounts of data, manual searches or programming may be used to verify the data, or even ETL or ELT (Extract, Load, Transform). When dealing with data volumes on the PB or even EB levels, however, these solutions become completely ineffective.

B. Data volume is tremendous, and it is difficult to judge data quality within a reasonable amount of time.

An amount of data that could be manipulated by characters increased every ten years during the industrial revolution. The quantity of data increased every three years after 1970. The quantity of data available on Earth may now be quadrupled every two years. In 2011, 1.8 ZB of new and copied data was generated worldwide. Gathering, cleaning, integrating, and ultimately obtaining the required high-quality data in a reasonable amount of time is challenging. Because big data contains a lot of unstructured information, cleaning it up and making it usable for further processing is a laborious operation. For data processing quality methods now in use, this presents a formidable obstacle.

C. Data change very fast and the "timeliness" of data is very short, which necessitates higher requirements for processing technology.

A "timeliness" of some data is quite brief since big data is evolving at a quick pace. Businesses risk getting inaccurate or out-of-date data if they can't handle their data demands over an extended period of time or get the necessary data in real time. Governments and businesses will make poor decisions as a result of processing and analysing this data, which will provide inaccurate or irrelevant results. There are currently very few commercially available real-time processing and analysis solutions for big data, and most of those that do exist are in the early stages of development.

D. No unified and approved data quality standards have been formed in China and abroad, and research on the data quality of big data has just begun.

An ISO 9000 standards were created in 1987 by the International Organisation for Standardisation (ISO) to ensure product quality and to provide advantages to businesses.

More than a hundred nations and regions throughout the globe are already putting these principles into practice.

Through its implementation, trade barriers are reduced and businesses in both local and foreign markets are able to better understand one another. Despite efforts to establish data quality standards starting in the 1990s, the ISO did not release the ISO 8000 standards until 2011 [16]. More than 20 nations have adopted this standard so far, although there are still plenty of disagreements. A flawless and mature set of criteria is required. Researchers in China and others have just recently started looking into the topic of big data quality, thus far, their findings are scant.

V. UNDERSTANDING QUALITY ASSURANCE FOR BIG DATA APPLICATION SYSTEM

QA in systems that manage big data applications is discussed in this section, along with its aim and methods. In addition, it addresses the main quality indicators together with contextual considerations. The following distinguishing characteristics of big data applications: (a) statistical analysis using massive datasets with several dimensions; (b) knowledge-based system evolution and ML; (c) intelligent decision-making under uncertainty; (d) non-oracle functions; and (e) complex visualisation. The demands, difficulties, and requirements related to quality assurance and QoS are made more intriguing by these distinctive aspects. Figure 2 displays the current research status on the subject of how to ensure the quality of big data application systems, which has lately attracted a lot of interest from engineers in Silicon Valley [17].

Figure 2: The Typical Types of Big Data Application Systems

A. *Scope and Process of Big Data Application Quality Assurance*

Big data applications use sophisticated cognitive algorithms and vast data sets to enable decision-making, prediction, and recommendation services. The common categories of big data applications are shown in Figure 2.

Various strategies are studied and implemented in QA for big data applications to guarantee that a big data system satisfies a set of quality metrics. One such approach of validating a large data application's quality is shown in Figure 3.

Figure 3: The Scope of Validation for Big Data Application System Quality

B. *Quality factors for big data application validation*

It is possible to apply traditional system quality metrics, such as security, resilience, and performance, to big data systems. They are mentioned in the list below.

System Performance – This metric shows how well the system is doing in terms of things like availability, response time, scalability, and overall performance.

- System Data Security – The security of a system that relies on big data might be assessed from several angles using this metric. There are a number of levels at which data security might be assessed using this parameter.
- System Reliability – The system's endurance in carrying out a particular function under certain circumstances for a given duration is assessed by this parameter.
- System Robustness - This metric measures how well a system can withstand perturbations without modifying its original, stable state.

VI. VALIDATION METHODS FOR BIG DATA APPLICATION

Software testing methodologies for evaluating big data applications in domains such as bioinformatics, data mining, intelligent systems, and learning-based activities are covered in this area.

A. *Program-based software testing*

Applications related to big data analytics first use program-based testing. A novel approach for methodically locating security holes in MapReduce applications and developing test cases is presented by Csallner et al. Symbolic program constraints are used to represent high-level MapReduce correctness standards, and the tested application is used to verify them. Shang et al. [18] created a method based on Handoop to examine the actions of the BDA App platform with the use of limited test data and large amounts of real-world data stored in the cloud.

B. *Classification based testing*

The two main components of a classification approach to software testing are 1. To identify errors in a portion of the results, train a classifier. 2. To find errors in the whole set of results, use the learnt classifier. A similar reference model is often utilised for classifier training.

Test oracle problems may be resolved, in particular, by pattern categorisation methods. Data mining may be used to enhance an old system's faulty specification. The unofficial input-output correlations of an outdated system are taught to classifiers via training. The classification of failure cases by Podgurski et al. [19]. The research does not examine how to differentiate between proper and unsuccessful program behaviours. Later, their research group suggests using classification trees to improve classifier outcomes. Bowring et al. develop a classifier using progressive machine learning on various software behaviours. The regression testing approach is used to test a series of small software modifications.

C. *Metamorphic Testing (MT)*

This method is often used to test programs without the use of Oracles. If a formal Oracle is unavailable or too expensive, we may encounter a test Oracle issue. The purpose of a test oracle is to help testers confirm the correctness of program output. When an oracle is unavailable, pseudo-oracles are used to verify test results. MT checks whether a program under test follows metamorphic relations, an anticipated set of traits. A metamorphic connection describes how a change in program input affects output. MT was used to evaluate scientific applications, including machine learning, bioinformatics, PDE, and image processing. MT detected problems in testing programs that special value testing is unable to detect while solving partial differential equations.

D. Learning-Based Testing

The process of implementing the many learning strategies and tools that facilitate testing of big data systems is included in this. Using learning-based testing (LBT), Meinke et al. created a method for automatically generating test cases for numerical applications. Before testing the actual software, the authors abstracted it using a polynomial model. The next step is to apply a satisfiability algorithm to the learnt model in order to produce the test cases.

E. Crowd-sourced testing

This method of testing makes use of a crowdsourcing community consisting of independent testers and/or hired engineers. Validating application systems based on machine learning, such a human face recognition system, is made easy and inexpensive using this technology. In recent times, crowdsourcing has been used in both the creation of mobile apps and mobile testing as a service. One reliable source is (uTest - The Professional Network for Testers). instance, uTest

F. Data model-based testing

Numerous data models may be used to facilitate the generation of test cases, since large data is what big data application systems rely on for their input values. This paper lays forth a method for automatically developing test cases for scientific programs that include several dependent input parameters Values, parameters, and the interactions between them make up the input data space, which is represented as a directed graph. The directed graph model allows for the automated generation of valid test cases. Using their approach, which follows the Markov chain probability rule, you may create test cases using a combination of random selection and weighting based on a likelihood of parameter values.

G. Rule-based software testing

Systems that rely on rules or knowledge might be tested using this method. Building test cases according to the rules defined by an expert system is the main concept. A rule-based approach to generating test data for Ada programs was suggested by Deason et al. in [20]. Their success demonstrated the practicality of creating test data using rules. Helping test engineers with test creation is where this document really shines. A method to test pattern creation based on VHDL-specific heuristic principles was provided by Andrews et al. [21]. The rule-based method improved test coverage, according to their data[22].

H. Conventional black-box software testing:

Big data application developers may rely on traditional black-box quality control methods to guarantee system performance and associated quality of service KPIs. Case studies, equivalence partitioning, decision tables, boundary value analyses, cause-and-effect graphs, use case testing, and many more are common examples.

VII. LITERATURE REVIEW

The implementation of various QA technologies to guarantee data application quality has been the subject of much research in the academic literature. There have been several articles that have provided general descriptions of data application quality issues.

This paper, Ji et al., (2020) reviewed and evaluated current quality assurance (QA) tools for fixing large data application quality problems. They searched key scientific databases to do a systematic literature review (SLR), which yielded 83 main and important papers on QA methods for big data applications. the data collected by each QA tool thus far. This study lays the groundwork for future research on large data application QA technologies and may assist developers in implementing appropriate QA strategies[11].

R. Kumar et al.,(2018) increasingly complicated nature of data analytical software products means that they demand more labour. The quantity and diversity of Data may be one of the causes. However, regardless of whether the testing is conducted using simulations or erroneous data, these software solutions always fail when evaluated in real life. Some quality requirements need to be defined in order to ensure that these software products function properly both before and after deployment. Better analytics software products may be produced in this manner [23].

Recently, U. Sivarajah,et al.,(2017) there has been a lot of interest from academics and practitioners in Big Data (BD) due to its ability to provide valuable insights for better decision-making. This is achieved by carefully reviewing and merging data sets. The purpose of this SLR is to analyse current and past trends in BDA in depth so that academics may better understand the area and its patterns. The following sections provide a synopsis of the information, an assessment of the contributions, and a list of the limits, consequences, and possible future study directions [24].

Therefore, (Zhang et al., 2017) It is crucial to do a comprehensive research on the topic of quality assurance in big data applications. In addition to introducing big data characteristics and quality traits, this article focusses on surveying methods for ensuring the quality of big data applications. Main topics covered include testing, MDA, monitoring, fault tolerance, verification, and prediction methods as means to guarantee the quality of big data systems. Moreover, the effects of big data features on big data uses are covered in this particular article[25].

In, (Kush, Hwang and Dattana, 2017) considers the quality assurance measures' impact on the efficiency of the AODV and DSR routing algorithms. The AODV and DSR protocols' performance differences are examined in various scenarios with varied stop times, speeds, and numbers of nodes using NS-2, the primary network simulator, NAM (Network Animator). The results are compared based on the scales given to the PDR[26].

As part of this paper, (Trinks and Felden, 2019 came up with this Image Mining program by using research methods from the field of design science. After that, the production prototype is given a quality class using a supervised ML method. The results showed that Image Mining may aid in QA and that there is a correlation between classification accuracy and latency[27].

(Clarke, 2017), a quality assurance structure is required to get certainty in these apps' results. This article outlines the difficulties and highlights the results of poorly planned and executed initiatives. It lays forth the fundamentals of risk assessment and management strategies, as well as potential areas for further study[28].

QA of big data applications is a vast area, with many associated difficulties and potential for future enhancement. Taking this into account, it is possible to indicate several crucial research gaps that are to be addressed despite the various recent developments in QA technologies. There is also one more troublesome aspect here, namely the absence of universally accepted quality standards which would guarantee that the software products used in data analysis will continue to operate as intended after being released into production. Furthermore, different QA techniques and tools have been proposed but their practical performance in practical application does not meet the expectations especially when facing the problems raised by big data issues of volume, variety and velocity. There is, thus, another gap with regards to how the QA processes are fitted into big data frameworks which may not have the right characteristics of handling big data by current means as they may not take into account the features of big data; for example, distributed; real time. In addition, and probably more importantly, there is a clear call for better quality assurance solutions that could forecast, identify and, subsequently, prevent the emergence of further quality problems. Feedback and QA related to big data characteristics and the ability and capacity of the QA tools and prospective landscapes of the big data applications are among crucial aspects that have been left out and should be considered more in the future for better and consistent practices of the big data applications.

VIII. CONCLUSION AND FUTURE WORK

Improved clinical decision-making via the use of technology may be possible with an use of big data and sophisticated analytics. To facilitate universal access and openness across healthcare organisations, however, the enormous volumes of data produced every year by the healthcare sector must be structured and partitioned.

In conclusion, the rapid expansion of big data applications necessitates robust and innovative QA methodologies to address the unique challenges posed by the 4V attributes. The heterogeneity, scale, and dynamic nature of big data require novel QA technologies that can ensure data accessibility, timeliness, credibility, and consistency. This paper has reviewed and analysed the principles and elements of QA, highlighting the importance of comprehensive and systematic approaches to validating big data systems. Various validation methods, including program-based, classification-based, and metamorphic testing, offer promising solutions for enhancing the quality and reliability of BDA. However, there remains a significant gap in standardised QA practices and technologies, particularly in real-time data processing and integration. To adapt to the ever-changing big data ecosystem, future studies should concentrate on creating consistent data quality standards and improving QA tools. By addressing these challenges, organisations can leverage high-quality big data applications to make informed decisions, enhance operational efficiency, and drive innovation.

REFERENCES

[1] C. L. Philip Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Inf. Sci. (Ny).*, 2014, doi: 10.1016/j.ins.2014.01.015.

[2] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, 2019, doi: 10.1016/j.cities.2019.01.032.

[3] C. Tao, J. Gao, and T. Wang, "Testing and Quality Validation for AI Software-Perspectives, Issues, and Practices," *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2937107.

[4] B. Jan *et al.*, "Deep learning in big data Analytics: A comparative study," *Comput. Electr. Eng.*, 2019, doi: 10.1016/j.compeleceng.2017.12.009.

[5] N. Laranjeiro, S. N. Soydemir, and J. Bernardino, "A Survey on Data Quality: Classifying Poor Data," in *Proceedings - 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing, PRDC 2015*, 2016. doi: 10.1109/PRDC.2015.41.

[6] I. Anagnostopoulos, S. Zeadally, and E. Exposito, "Handling big data: research challenges and future directions," *J. Supercomput.*, 2016, doi: 10.1007/s11227-016-1677-z.

[7] J. Gao, C. Xie, and C. Tao, "Big data validation and quality assurance - Issuses, challenges, and needs," in *Proceedings - 2016 IEEE Symposium on Service-Oriented System Engineering, SOSE 2016*, 2016. doi: 10.1109/SOSE.2016.63.

[8] S. T. Lai and F. Y. Leu, "Data preprocessing quality management procedure for improving big data applications efficiency and practicality," in *Lecture Notes on Data Engineering and Communications Technologies*, 2017. doi: 10.1007/978-3-319-49106-6_73.

[9] N. Garg, S. Singla, and S. Jangra, "Challenges and Techniques for Testing of Big Data," in *Procedia Computer Science*, 2016. doi: 10.1016/j.procs.2016.05.285.

[10] R. T. Kaleem Ullah, Majid Ali Tunio, Zahid Ullah, Muhammad Talha Ejaz, Muhammad Junaid Anwar, Muhammad Ahsan, "Ancillary services from wind and solar energy in modern power grids: A comprehensive review and simulation study," *J. Renew. Sustain. Energy*, vol. 16, no. 3.

[11] S. Ji, Q. Li, W. Cao, P. Zhang, and H. Muccini, "Quality assurance technologies of big data applications: A systematic literature review," *Appl. Sci.*, 2020, doi: 10.3390/app10228052.

[12] B. Uma shankari, "Quality assurance in health care systems with special reference to hospitals in Coimbatore city," 2013.

[13] A. Syed, "Dental sensitivity associated with consumption of fizzy drinks: A cross sectional study," *Pakistan J. Med. Dent.*, 2018.

[14] J. Thomas, "The Effect and Challenges of the Internet of Things (IoT) on the Management of Supply Chains," vol. 8, no. 3, pp. 874–878, 2021.

[15] D. McGilvray, *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)*. 2008.

[16] Q. Wang, Jun-Ling and Li, Hun and Wang, "Research on ISO 8000 series standards for data quality," *Stand. Sci.*, pp. 44--46, 2010.

[17] C. Tao and J. Gao, "Quality assurance for big data application-issues, challenges, and needs," in *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE*, 2016. doi: 10.18293/SEKE2016-166.

[18] W. Shang, Z. M. Jiang, H. Hemmati, B. Adams, A. E. Hassan, and P. Martin, "Assisting developers of big data analytics applications when deploying on Hadoop clouds," in *Proceedings - International Conference on Software Engineering*, 2013. doi: 10.1109/ICSE.2013.6606586.

[19] A. Podgurski *et al.*, "Automated support for classifying software failure reports," in *Proceedings - International Conference on Software Engineering*, 2003. doi: 10.1109/icse.2003.1201224.

[20] D. B. Brown, K. H. Chang, and J. H. Cross, "A Rule-Based Software Test Data Generator," *IEEE Trans. Knowl. Data Eng.*, 1991, doi: 10.1109/69.75894.

[21] A. Andrews, A. O'Fallon, and T. Chen, "A Rule-Based Software Testing Method for VHDL Models," in *Proc. IFIP International Conf. of VLSI …*, 2003.

[22] V. Rohilla, S. Arora, P. S. Nirwan, and V. Purohit, "Recommendation System Using Location-Based Services," *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3833939.

[23] R. Kumar, B. Subhash, M. Fatima, and W. Mahmood, "Quality assurance for data analytics," *Int. J. Adv. Comput. Sci. Appl.*, 2018, doi: 10.14569/ijacsa.2018.090821.

[24] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, 2017, doi: 10.1016/j.jbusres.2016.08.001.

[25] P. Zhang, X. Zhou, W. Li, and J. Gao, "A survey on quality assurance techniques for big data applications," in *Proceedings - 3rd IEEE International Conference on Big Data Computing Service and Applications, BigDataService 2017*, 2017. doi: 10.1109/BigDataService.2017.42.

[26] A. Kush, C. J. Hwang, and V. Dattana, "Big data analytics on MANET routing standardization using quality assurance metrics," in *FTC 2016 - Proceedings of Future Technologies Conference*, 2017. doi: 10.1109/FTC.2016.7821610.

[27] S. Trinks and C. Felden, "Image Mining for Real Time Quality Assurance in Rapid Prototyping," in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, 2019. doi: 10.1109/BigData47090.2019.9005514.

[28] R. Clarke, "Quality assurance for security applications of big data," in *Proceedings - 2016 European Intelligence and Security Informatics Conference, EISIC 2016*, 2017. doi: 10.1109/EISIC.2016.010.