2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 979-8-3503-9926-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACITE57410.2023.10182930

Implementation of smart future cities with big data analytics integrated with cloud computing

Molli Srinivasa Rao
Department of CSE,
Dadi Institute of Engineering & Technology
Anakapalle, Visakhapatnam, India
drmollisrinivasarao@gmail.com

Lovi Raj Gupta
Lovely Professional University,
Phagwara, Punjab, India
lovirajgupta2323lpu@gmail.com

Rajat Singh
Uttaranchal Institute of Technology
Uttaranchal University
Dehradun, Uttarakhand, India
rajat@uumail.in

Litan Debnath
Department of Civil Engineering,
School of Engineering,
SR University,
Warangal, Telangana, India
litan.debnath@sru.edu.in

Patel Chaitali Mohanbhai
Department of Faculty of Computer Science,
Shri C.J. Patel College of Computer Studies,
Sankalchand Patel University
Visnagar, Gujarat, India
cmpatel.fcs@spu.ac.in

N. Balaraman
Department of CSE,
St. Martin's Engineering College
Secunderabad Telangana, India
padmabala.raj@gmail.com

Abstract: **Many cities and towns have recently formed aspirations to become smart cities. These networks are attempting to implement the smart city concept in order to better increase supportability, inhabitant personal satisfaction, and financial outcomes. The concept of a "smart city" focuses on utilising cutting-edge technologies to improve several administrations and endeavours, including transportation, energy, healthcare, and education, as well as to more effectively utilise the city's resources and foster new economic opportunities.. One of the promising instruments to help with these endeavors is big data innovation. Smart cities can make better decisions by using big data that has been amassed through time in many different businesses. This study distinguishes the various dynamic methods used in smart cities. This design empowers the incorporation of different advances, for example, fog registering and distributed computing, to help the various scientific and decision-making processes expected to completely use the big data right now accessible. Numerous big data applications can thusly involve these administrations in the smart city and use them for different targets.**

Keywords: Big data, cloud computing, middleware,smart city.

I. Introduction

As enabling technologies evolve and become more accessible, smart cities are growing swiftly. Technologies like the CPS, WNS, and IOT, which rely on connectable items, sensors, actuators, and controllers, have enabled several applications for "smart cities." These technologies all the while produced colossal measures of data about their workplaces, functional status, frameworks, and occupants [1]. The massive volume of data that was acquired, in addition to its high velocity and variety, produced Big Data [2]. One of the key IoT platforms is the cloud, which has also given rise to a number of platforms for smart cities, such as the Cloud of Things [3], CPS [4], WSN [5,], and others. The size of big data sets has increased over time, requiring the creation of sophisticated and efficient methods for handling, analyzing, and utilizing them. The end goal is to create autonomous decision-making tools for smart cities that can make use of this vast amount of data. The necessary processing and storage facilities to support these solutions are provided by large computing infrastructures, such as the cloud. When working on ongoing projects like traffic control, crowd control, and catastrophe recovery, quick judgements must be made [6]. The development of numerous models and procedures that can make use of the cloud while also offering a constant, adaptable, and variety of elements is essential [7].

One proposed solution to this problem is to relocate the handling and storage procedure closer to the actual weather conditions [8] utilising fog or edge nodes. These have varying degrees of handling and storage constraints and can deliver faster, more space-efficient, and more versatile reactions [9]. However, fog nodes by themselves are unable to process as much massive data as cloud nodes. Therefore, it is crucial to design original strategies for fusing all of these technologies. For the applications associated with smart cities, the incorporation will give a more complete arrangement that can use these frameworks' and stages' best highlights.

This study distinguishes the different dynamic cycles utilized in smart cities as well as the significant data analytics levels engaged with each kind. The foundation of this strategy is the construction of big data management methods, tools, operations, and analytics techniques as services. In the big data space, specifically, certain data gathering, association, planning, moving, examination, and revealing strategies can be sent as administrations. All administrations don't need to be presented on a similar stage. Now, any service can be positioned where it will perform at its peak and benefit the application consuming it the most.

II. Related Work

A. A. Smart Cities

A smart city strives to maximize resource usage, increase sustainability, and minimize environmental harm while simultaneously enhancing people's quality of life. For a multitude of reasons, including the rate of global population expansion, resource scarcity, the developing spotlight on maintainability and energy proficiency, and the requirement for more grounded economies, numerous urban communities all through the world are taking on the "smart city" thought.

This encompasses all of the fundamental city infrastructures, such as the transportation network, the buildings, the communications network, and the people that live there [10]. Smart city management, funding, regulation, and planning require a strong framework. Four levels make up one likely design: city markers, city desires, city parts, and city content [11]. It's likewise fundamental to distinguish the key elements influencing smart city projects. These variables incorporate, however, are not restricted to, the executives and association, innovation, administration, strategy setting, economy, ICT framework, individuals and networks, and the common habitat [12]. To achieve the goal of the smart city, various emerging and modern developments are also crucial.

B. B. Big Data Analytics and Smart Cities

Big data sets are typically gathered over time and provide a significant pool from which data can be retrieved to enable sophisticated and improved decision-making algorithms. These data collections are enormous, much of the time unstructured, dynamic, and comprised of a few sorts. As a result, conventional ICT infrastructures often cannot handle the processing, administration, and storage of vast amounts of data [13]. Cloud computing is a useful platform for managing various big data applications kinds. To more readily uphold big data management and investigation, a cloud-based design named data as-a-administration has arisen [14].

These constantly gather information and move it into the ICT framework, producing huge information simultaneously. Big data investigation methods that are both reasonable and productive are turning out to be increasingly more significant for utilizing this information effectively. Big data analytics creates intelligent data that enhances various decision-making processes by using techniques including sorting, cleaning, transforming, and modelling exponentially rising amounts of data [15]. By enabling data-based decision-making, big data analytics may be utilised to enhance the services provided by smart cities. Large information examination applications in smart cities include those that enable intelligent frameworks, intelligent grid infrastructure, intelligent grid education, intelligent traffic signals and foundation, intelligent structures and frameworks, intelligent medical care, and smart management and organisation [7].

C. C. Technologies for Smart Cities

WSN an organization of organized sensors used to screen unequivocal circumstances and city-state information is one of these forward leaps [16] [17]. Since sensors efficiently screen, record, and transmit information to capacity and handling centers. IoT is an essential piece of technology that underpins the idea of the smart city. It is a system that links tangible objects, including autos, personal electronic devices, actuators, and sensors, and enables the communication between them. IoT enables the smart city to integrate devices from diverse industries and domains to create useful applications [18]. The number and variety of big data sources are growing, and the Internet of Things is just one of them. On an adaptable, resource-rich platform enabled by cloud computing, big data solutions for smart cities can be produced and operated. Better access to data collected, powerful handling capabilities, and more precise IoT control (creating CoT) were made possible by the integration of the cloud with IoT. A fog node functions similarly to cloud nodes and can provide a variety of administrations, including capacity, calculation, security, control, and area of administration [19]. Fog nodes can help, but their size and strength may have a limit on how much.

D. D. Companion Work

Many of the applications for large data experimentation share traits with more complicated ones. They do, however, also possess a few characteristics that are mandated by the nature of the massive amounts of data that they are meant to process. These include the creation and collection of data, its management and organization, its storage and movement, the methods of analysis, and the level of accuracy of the results. They, therefore, impose particular restrictions when they are developed and put into use [20] [21]. Many individuals researched the expected advantages of big data investigation for smart city administrations [22]. Presents two examples of when IoT devices put in diverse smart city sectors generated enormous amounts of data, illustrating the value of analytics in improving decisions [7] and gave a survey of a few potential big data examination applications.

III. Methodology And Implimentation

A. Architecture

The various forms of innovation that go into creating a smart city provide different kinds of help to gather, channel, store, process, and decipher information to expand metropolitan services for people and resource utilization. Generally, a ton of these administrations need to settle on choices given the information assembled. We talk about their attributes, applications, and reasonableness for different kinds and levels of data investigation.

1) Management Options

This kind of dynamic offers immediate, ongoing, intelligent, and shut-circle controls for the administrations of smart cities. These decisions, which need to be made right away, are based on the most current situations that have been observed as well as planned configurations or control rules. To achieve the set objectives, these decisions primarily serve as timely instructions to be applied to the management of a few services and resource aspects in smart cities. For instance, a set of control rules might be employed to coordinate a few smart traffic signals at connected intersections in a city to offer the optimal traffic flow [23].

2) Selection of Configurations

Deciding how a specific help in a smart city can be created or changed to accomplish the most ideal presentation given the actual assets and limitations accessible is the central objective of setup choices. The reconfiguration procedure relies on extensive analysis, enhanced revelation, new rule age, advancements, assessments, gauges, and the capacity to make new control leads or alter existing ones.

3) Planning Selections

Long-term planning processes are required when making choices concerning the future augmentations, enhancements, reconstructions, and reconfigurations of the infrastructure elements of smart cities. For instance, choosing which roads to extend or fix first to decrease traffic issues, or assessing the effects of building a tall skyscraper at a specific location on the environment and nearby infrastructure [24].

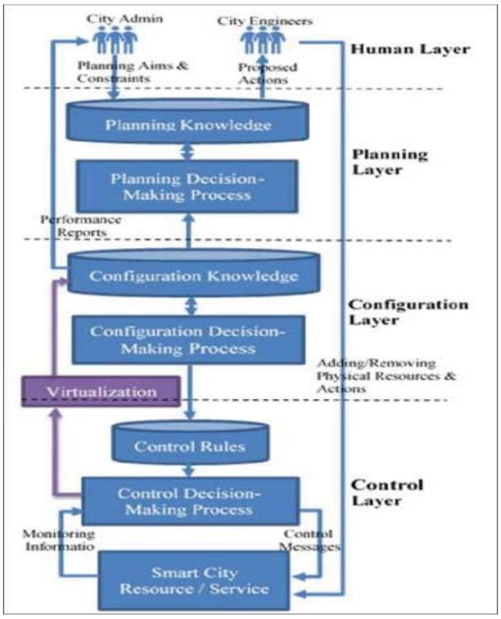

Fig. 1. levels of decision-making in a smart city

Going with planning decisions requires the improvement of an exhaustive arranging knowledgebase framework integrating all got information from the total city foundation, both current and verifiable.

4) The Big Picture of Every Type of Decision

The three principal kinds of decisions are all made in a smart city. Every classification is related to and influences others in various ways. Figure 1 shows a broad overview of the relationships between the various decision sorting. The four tiers that are visible are, the arrangement layer, the arrangement layer, the control layer and the human layer. City planners and architects concentrate on the human layer. The nearby authorities will lay out the proper execution estimations, arranging goals, and imperatives utilizing the spending plan and current regulation.

B. Implementation of prototype

For the prototype implementation, the Quality of Life (QoL) data, which contains a number of indicators to measure the QoL, including Culture and Leisure, Economy and Employment, Crime and Safety etc., was chosen. A survey designed to get input from residents on the pertinent indication for their location is used to evaluate each indicator. As shown in Figure 2, the dataset has values for the questionnaire replies spanning several years that have been aggregated. The data had a size of roughly 0.7 MB.

1) Application architecture prototype

A simple prototype application was made using Hadoop and Spark to demonstrate the suggested huge data analysis architecture's practical usefulness. The Bristol City Council uses survey questionnaires to find out what the public thinks about certain metrics. However, none of the different indications have an implicitly quantitative measure. When evaluating Bristol's Quality of Life, decision-makers may find it useful to use such a measurable indicator.

We suggest arranging the available Bristol open data in a hierarchical manner in order to generate the aforementioned quantitative measurements. We divided the data for every indicator during preprocessing into year categories to allow parallel processing using the MapReduce framework. Every year for which data were available, the survey results were already sorted by wards. The Hadoop framework automatically divides the data in a MapReduce-based application into worker units called Mappers in a stage referred to as Mapping. The data created by the Mappers is provided to the Reducers after some scrambled and sorted before the Mappers finish their planned jobs. The Reducers must then transform the data into information that is useful.

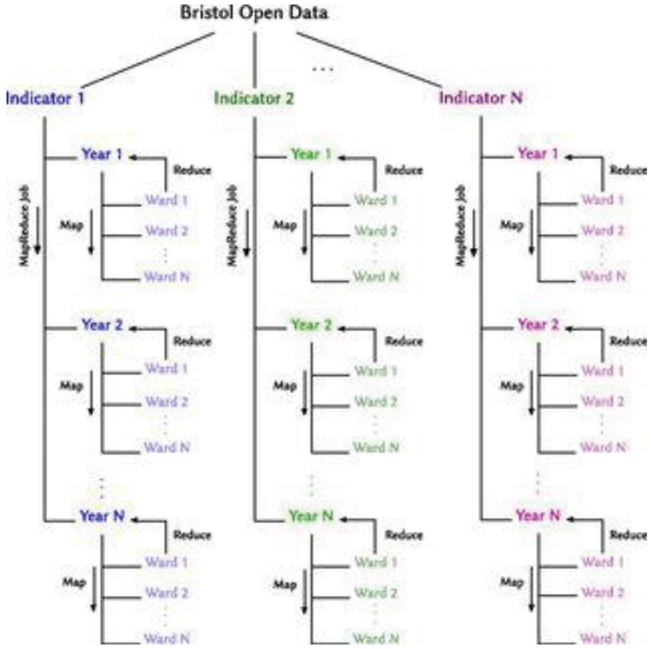

Fig. 2. Prototype architecture

IV. Results And Discussion

1) A. Implementation of Hadoop

For this implementation, 386 mappers and 4 reducers in total were created. It took about 16 seconds to complete each task. The tests were carried out:

TABLE I. SPARK EXECUTION OUTCOMES

Mode of Execution	Computation nodes	Eexecution time per task	Execution time (total)
cluster	1	16 sec	26 min
local	1	0.02 sec	6 sec
cluster	3	16 sec	13 min

Based on the given data, the computed Pearson's correlation between the two indicators was 0.2406. The most common metric for statistically comparing two datasets is this one. This parameter's value is always in the range of 1 and -1. One denotes a significant negative correlation, zero denotes no association, and 0 denotes no correlation at all. The correlation between the two variables in question is weak, as indicated by the value of 0.2406. The positive figure shows that there is some improvement in the city's crime and safety condition as a result of better economic and employment prospects. Additionally, it is intuitively possible to construct a causal link between the two indications. Figure 3 makes it possible to visually verify this correlation. Despite the fact that the broad trends are correlated, there are still some cases where the connection is poor. For instance, there is a significant fall in the economy and employment chances from 2006 and 2007. The crime and safety situation, however, slightly improves. Then, there aren't many differences between 2009 and 2010 in terms of the economy or employment chances. However, the situation with regard to crime and safety has improved a little. The only overall weak link is these little variations in trends. However, since the sample set is too tiny to be statistically reliable, care should be taken when drawing conclusions.

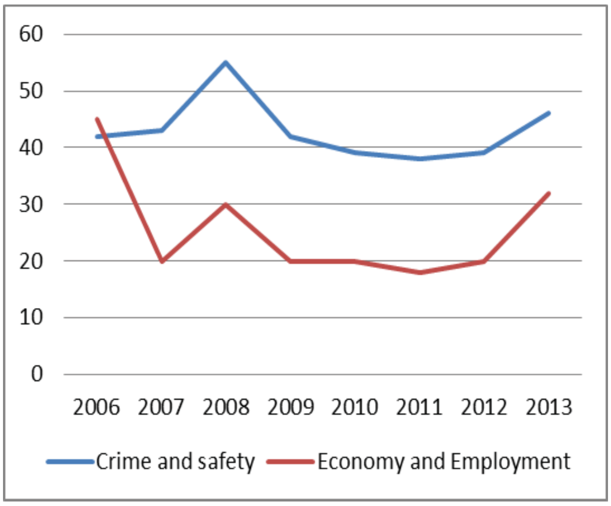

Fig. 3. Citizens' perceptions of crime and safety in relation to the economy and employment

V. Conclusion

Smart cities provide the opportunity to connect people and places while enhancing urban planning and management through the use of cutting-edge technologies. The vast volumes of data that are generated every minute in an urban setting as a result of anthropogenic, natural environmental occurrences, socioeconomic, or other activities are the foundation of smart cities. These data must be collected, processed, analysed, and presented. Data for smart cities can be promptly gathered from a variety of sensors, smart phones, and individuals and combined (or linked) with city data repositories in order to use analytical reasoning and provide the data required, or fresh information for decision-making for improved urban administration. Now that information and communication technology has advanced, it is possible to organise, understand, and provide people with access to current and crucial information.

In this paper, we examined big data analytics for cloud-based smart cities. Data collection, preparation, semantic linkage, and the use of proper data mining, machine learning, or statistical analysis methodologies are a few of the factors that need to be carefully considered. Additionally, interaction with subject matter experts is required to uncover the fundamental connections and dependencies between various data pieces due to the multidisciplinary character of the application domains for smart cities. The suggested architecture gives you the fundamental components you need to construct the functionality you need for a cloud-based big data analytical application that incorporates data from smart cities. We created a MapReduce prototype as a proof of concept to show how cloud infrastructure could be used to evaluate various samples of Bristol Open Data. Comparisons are made between the results of the prototype's implementation using Spark and Hadoop. The findings demonstrate that when jobs are sent to the cluster, Hadoop incurs significant overhead, most likely as a result of costly data access procedures. In terms of speed and overhead, Spark performs far better. Therefore, Spark is a preferable option for the chosen Bristol open dataset.

Due to its aggregated nature, the dataset that may be accessible through the Bristol Open Data website technically does not fit the criteria of Big Data. The proof-of-concept, however, demonstrates how Big Data applications may be used with similar IT infrastructures. Based on the study's findings, we examined how effectively the flexibility of the cloud might meet the demand for data analytics in smart cities. The successful use of the prototype shows the value of cloud-based platforms for data analytics in smart cities.

Since the majority of local administration data are not in the public domain, open data were used in the prototype application. On the one hand, several cities offer monthly or even annual summaries of the daily data, greatly reducing the total size/volume of such data. Although the aggregated data lacks more precise geo-coordinates/locations, the total dimensions are considerable. As a result, the privacy of users is protected and managing such data is made simple. However, only a tiny sample set is made available, making it difficult to gain deeper understanding and pinpoint specific correlations between data points. Although this information tends to grow with time, it might not be as extensive and detailed as it is for other major cities. The open data portal for Bristol hosts these collected data sets, which range in size from a few hundred K to Megabytes.

Our forthcoming research will expand the technical infrastructure to seek for connections among other indicators in the currently accessible open data as well as explore the application of semantic sources like RDF stores. The goal of this effort would be to pinpoint technological implications and constraints while also offering workable solutions.

References

[1] R. Kitchin, "The real-time city? Big data and smart urbanism," GeoJournal, vol. 79, pp. 1-14, 2014.

[2] J. Dutcher, "What Is Big Data?" datascience@ Berkeley, September 03, 2014. [Online]. Available: https://datascience.berkeley.edu/what-is-big-data/. [Accessed: Jan. 11, 2023].

[3] N. Mohamed, S. Lazarova-Molnar, and J. Al-Jaroodi, "Cloud of things: Optimizing smart city services," in 2017 7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO), 2017, pp. 1-5.

[4] L. Gurgen, O. Gunalp, Y. Benazzouz, and M. Gallissot, "Self-aware cyber-physical systems and applications in smart buildings and cities," in 2013 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013, pp. 1149-1154.

[5] T. Watteyne and K. S. Pister, "Smarter cities through standards-based wireless sensor networks," IBM Journal of Research and Development, vol. 55, no. 1.2, pp. 7-1, 2011.

[6] N. Mohamed and J. Al-Jaroodi, "Real-time big data analytics: Applications and challenges," in 2014 International Conference on High-Performance Computing & Simulation (HPCS), 2014, pp. 305-310.

[7] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, and J. Al-Jaroodi, "Applications of big data to smart cities," Journal of Internet Services and Applications, vol. 6, no. 1, pp. 1-15, 2015.

[8] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, 2012, pp. 13-16.

[9] M. Aazam and E. N. Huh, "Fog computing and smart gateway-based communication for the cloud of things," in 2014 International Conference on the Future Internet of Things and Cloud, 2014, pp. 464-470.

[10] V. Panwar, D. K. Sharma, K. V. P. Kumar, A. Jain, and C. Thakar, "Experimental Investigations And Optimization Of Surface Roughness

In Turning Of EN 36 Alloy Steel Using Response Surface Methodology And Genetic Algorithm," Materials Today: Proceedings, 2021. [Online]. Available: https://doi.org/10.1016/j.matpr.2021.03.642. [Accessed: Jan. 11, 2023].

[11] A. Jain and A. K. Pandey, "Modeling and Optimizing of Different Quality Characteristics in Electrical Discharge Drilling of Titanium Alloy (Grade-5) Sheet," Material Today Proceedings, vol. 18, pp. 182–191, 2019. DOI: 10.1016/j.matpr.2019.06.292.

[12] H. Chourabi et al., "Understanding Smart Cities: An Integrative Framework," in 2012 45th Hawaii International Conference on System Sciences, 2012, pp. 2289–2297. DOI: 10.1109/HICSS.2012.615.

[13] J. S. Ward and A. Barker, "Undefined by Data: A Survey of Big Data Definitions," arXiv preprint arXiv:1309.5821, 2013.

[14] Z. Zheng, J. Zhu, and M. R. Lyu, "Service-Generated Big Data and Big Data-as-a-Service: An Overview," in 2013 IEEE International Congress on Big Data, 2013, pp. 403–410. DOI: 10.1109/BigData.Congress.2013.85.

[15] S. LaValle et al., "Big Data, Analytics, and the Path from Insights to Value," MIT Sloan Management Review, vol. 52, no. 2, pp. 21–32, 2011.

[16] A. Jain, A. K. Yadav, and Y. Shrivastava, "Modelling and Optimization of Different Quality Characteristics in Electric Discharge Drilling of Titanium Alloy Sheet," Material Today Proceedings, vol. 21, pp. 1680–1684, 2019. DOI: 10.1016/j.matpr.2019.12.010.

[17] T. Watteyne and K. S. Pister, "Smarter Cities through Standards-Based Wireless Sensor Networks," IBM Journal of Research and Development, vol. 55, no. 1.2, pp. 7:1–7:11, 2011. DOI: 10.1147/JRD.2010.2091538.

[18] A. Zanella et al., "Internet of Things for Smart Cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22–32, 2014. DOI: 10.1109/JIOT.2013.2296513.

[19] A. Giordano, G. Spezzano, and A. Vinci, "Smart Agents and Fog Computing for Smart City Applications," in International Conference on Smart Cities, 2016, pp. 137–146. DOI: 10.1007/978-3-319-44447-5_12.

[20] J. Al-Jaroodi and N. Mohamed, "Characteristics and Requirements of Big Data Analytics Applications," in 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC), 2016, pp. 426–432. DOI: 10.1109/CIC.2016.86.

[21] D. Sharma, V. Kudva, V. Patil, A. Kudva, and R. S. Bhat, "A convolutional neural network based deep learning algorithm for identification of oral precancerous and cancerous lesion and differentiation from normal mucosa: A retrospective study," Engineered Science, vol. 18, p. 278-287, 2022.

[22] S. Kumar, "Reviewing Software Testing Models and Optimization Techniques: An Analysis of Efficiency and Advancement Needs", J. Comp. Mech. and Mgmt., vol. 2, no. 1, pp. 32–46, Feb. 2023.

[23] M. Shah, N. Naik, B. K. Somani and B. Z. Hameed, "Artificial intelligence (AI) in Urology-Current use and future directions: An iTRUE study," in Turkish Journal of Urology, vol. 46, no. Suppl 1, pp. S27-S32, Nov. 2020, doi: 10.5152/tud.2020.20291.

[24] S. Deepa, "Metaheuristics for Multi Criteria Test Case Prioritization for Regression Testing", J. Comp. Mech. and Mgmt., vol. 1, no. 1, pp. 42–51, Oct. 2022.