A Framework for Smart City Traffic Management utilizing BDA and IoT

Jayalakshmi Nagalapuram
Department of Computer Science and Engineering, PRIST University, Thanjavur, India
jayanaga1984@gmail.com (corresponding author)

S. Samundeeswari
Vandayar Engineering College, Thanjavur, India
samundeeswari172@gmail.com (corresponding author)

Received: 3 June 2024 | Revised: 18 June 2024 | Accepted: 20 June 2024

ABSTRACT

This paper explores a new approach to traffic flow optimization in smart cities, harnessing the combined power of Big Data Analytics (BDA) and the Internet of Things (IoT). The system utilizes a citywide network of connected sensors to acquire live traffic information, including vehicle speeds, density, and congestion points. These data are thereafter processed applying some top-notch BDA algorithms to identify traffic anomalies and forecast congestion levels, and generate actionable insights. By analyzing this information, the system can dynamically adjust traffic signals, recommend alternative routes, and improve traffic efficiency in real-time. The system's adaptive learning capabilities allow it to continuously enhance its predictions based on new data, ensuring its effectiveness in managing evolving traffic patterns. This intelligent traffic management solution promises to significantly reduce congestion, ameliorate overall mobility and road safety, and contribute to a more sustainable city environment.

Keywords-big data;BDA; IoT; industrial internet

I. INTRODUCTION

Internet of Things (IoT) refers to a collection of physical devices that are connected and equipped with sensors. This includes a broad range of interconnected devices, such as mobile phones, wearable technology, commercial equipment, and a variety of electronic devices, in addition to autonomous motors and home appliances [1, 2]. Anticipating the future surroundings of website users is crucial in order to mitigate traffic congestion and problems related to it. By adopting device learning techniques, site visitor forecasting may be done information-driven, using historical visitor data to make precise forecasts. For authority groups, these forecasts are essential because they allow them to promptly decide how to regulate traffic based on information obtained from site visitors in conjunction with flow forecasting [3–8]. There are two types of traffic congestion prediction: short-range and long-range forecasting. Utilizing both historical and real-time traffic data, short-range forecasting focuses on projecting traffic conditions for short-term events, ranging from hours to seconds, which helps customers better plan their travel itineraries [9, 10]. On the other hand, long-range forecasting forecasts the patterns of site visitors for entire future days. Businesses can maximize operations by using this information to grow their sites' visitor management and signal control programs [11].

The prediction of visitor flow in metropolitan areas is often dependent on data from the site's visitor sensors, which tally the number of cars that pass through. However, due to financial and practical limitations, installing sensors on every street is not feasible. Many procedures use data from a subset of roadways to anticipate gliding on both monitored and unmonitored roads in order to get around this problem. These procedures look at the temporal and spatial correlations between sensor data, identifying patterns and correlations in nearby and far-off locations [12, 13]. For instance, several trends incorporate more information sources than just traffic flow, effectively forecasting drift styles in metropolitan areas similar to those where data are collected. In urban places, traffic congestion is a major nuisance. This research presents a dynamic avenue visitor sign manipulator that optimizes traffic glide by using Artificial Intelligence (AI). This gadget can significantly minimize traffic congestion and enhance global traffic management by dynamically modifying signal timings based entirely on real-time site visitor conditions [14].

II. LITERATURE REVIEW

Authors in [15] systematically map scholarly works exploring AI-powered learning systems, analyzing 147 studies published between 2014 and 2020. Their purpose is to identify key interventions, research themes, analytical methods, and future directions for designing AI-powered learning systems that address precise learning objectives and enhance user experiences. The domain of Big Data Analytics (BDA) has experienced substantial progress based on a variety of research topics and agendas, including design, behavioral, and economic foci. Additionally, it is discussed that there is a need to clarify some of the elements that shape and structure the debate within the IoT field and its associated ecosystem, both of which are still in their infancy and particularly focused on technology. Moreover, conceptual transparency is considered a prerequisite for the effective integration of ideas across multiple disciplines [16]. In this section, the term "discipline" is used conceptually, rather than in a way to represent some specialization similar to big data or IoT analytics, aiming to refer to some emerging groups within the industry. On the other hand, the following claims include this characterization of IoT devices, which represents a particular challenge and focuses in many ways on leadership skills in the field of big data analysis and advertising [17].

Intelligent Transportation Systems (ITSs) are quickly developing to meet the increasing demand for more sustainable, safer, and more efficient modes of transportation. These applications include autonomous vehicles, mobility prediction, and site visitor control. Authors in [18] examine important ITS additives, such as Intelligent Traffic Lights, Mobility Prediction, and Vehicular Ad-hoc Networks, and analyzes their development, security requirements, and real-world applications in building eco-friendly smart cities. The study also emphasizes the vital role ITS play in promoting sustainable smart cities, including case studies and proactive measures that automakers support to guarantee mutual benefits and interoperability. Modern investigative research relies heavily on multimedia technologies, leading to societies becoming increasingly dependent on them [19]. To optimize smart community systems, efficiently analyzing big data collected from these communities is crucial. Several methods have been proposed for processing these data using Hadoop mechanisms [20-24]. While existing approaches like CiDAP offer valuable solutions, they often lack accessibility for further research and may not provide the most efficient processing mechanisms, highlighting the need for more comprehensive and accessible solutions for managing the vast amount of data generated by IoT technologies [23]. Figure 1 represents the process of smart city traffic management.

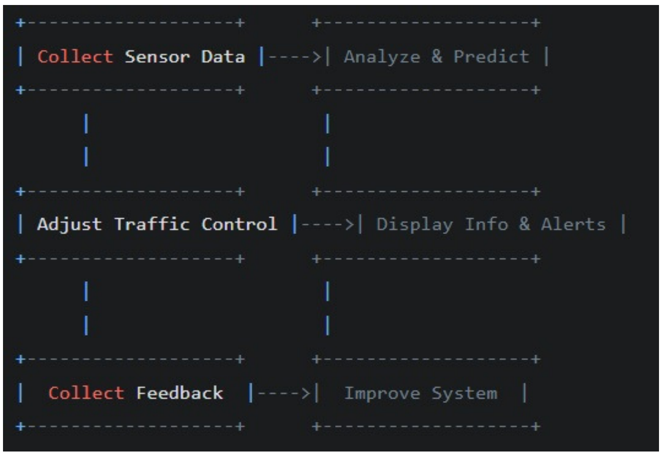

Fig. 1. Process of smart city traffic management.

Currently, no single model can effectively handle the diverse data generated by IoT devices in smart communities. This lack of a universal approach necessitates a comprehensive architecture capable of collecting, processing, and analyzing data from various sources. Existing research highlights key challenges such as data collection. Smart community development relies on noise removal, analytics, and decision-making. To overcome these challenges, a specific architecture based on distributed and parallel processing is proposed, leveraging multimedia BDA to enhance storage and processing efficiency within the IoT-enabled infrastructure.

III. PROPOSED WORK

The proposed architecture employs a parallel and distributed approach to handle multimedia big data sources efficiently, encompassing weather-, water-, traffic-, and health-related data along with online traffic video processing information. This architecture, illustrated in Figure 2, involves data gathering from various devices, including environmental, security, power, facility, and transportation monitoring sensors. The collected data, representing diverse multimedia information, are then curated and managed through a distributed caching system. This process, known as edge caching, involves collecting, aggregating, and storing data from distinct applications, such as smart health monitoring, traffic management, traffic control, and weather monitoring. The pre-processed data are then fed into the distributed and parallel computational architecture for processing using the proposed model rules.

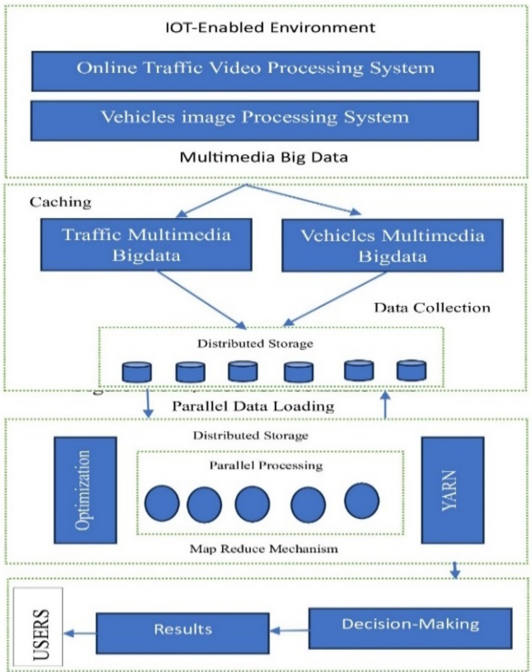

Fig. 2. Proposed service framework.

A. *Data Processing and Analysis for Big Data*

The core of this architecture lies in the Data Processing and Analysis for Big Data, while managing the actual processing of data. The principle behind this phase is parallel processing, where multiple tasks are executed concurrently to save time. By distributing workloads across multiple processors, parallel processing leverages the power of multiple CPUs to achieve faster execution. These data are processed using robust streaming platforms and advanced machine learning algorithms, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks.

B. Load Balancing and Parallelism

Load balancing is an essential part of parallel processing. It ensures that workloads are distributed evenly across the available computing processors, optimizing resource utilization and minimizing processing time. This balancing act aims to minimize processor usage, reduce response times, maximize throughput, and prevent overload. The Data Processing and Analysis for Big Data phase first implements a load balancing algorithm to divide the data load evenly across all servers. This equitable distribution optimizes system effectiveness, allowing each server to process an identical portion of the data and generate output simultaneously. The Data Processing and Analysis for Big Data phase first implements a load balancing algorithm to divide the data load evenly across all servers. This equitable distribution optimizes system effectiveness, allowing each server to process an identical portion of the data and generate output simultaneously.

C. MapReduce and Hadoop Ecosystem

For efficient data processing, the system utilizes the MapReduce programming mechanism. This phase leverages the powerful combination of MapReduce and Hadoop Distributed File System (HDFS). While HDFS is the primary storage system, alternative database solutions like HBASE and HIVE, both SQL-based, are also available for managing historical data (either offline or in-memory).

D. Bridging the Gap: Real-time Data Processing with Apache Spark

A dedicated device for real-time processing is required for both offline review and real-time information streaming. This gap is filled by Apache Spark, a powerful, multipurpose engine for handling enormous amounts of data that are easily integrated in the Hadoop environment. Spark enables processing big data quickly and allows reusing code across streaming applications. Its distributed processing framework, which is open-source, achieves exceptional speed by utilizing efficient execution and in-memory caching. In addition, the process of creating and managing Spark clusters is made easier by Amazon's built-in support for Apache Spark on Hadoop Yet Another Resource Negotiator (YARN).

E. Big Data Service Administration

The Big Data Service Administration phase forms the lowest level of the architecture, responsible for communication with third-party interfaces (including humans and objects). This phase serves as a crucial intermediary between the system and the end user, facilitating intelligent decision-making and event communication to relevant citizens through designated centers. The Big Data Service Administration phase forms the lowest level of the architecture, responsible for communication with third-party interfaces. This phase serves as a crucial intermediary between the system and the end user, facilitating intelligent decision-making and event communication to relevant citizens through designated centers. It plays a key role in generating insightful conclusions and disseminating relevant information, ensuring a smooth flow of data and actionable insights. Moreover, its flexibility allows for independent implementation, integration with other systems, or deployment as a network interface, enhancing its versatility and adaptability within diverse environments.

F. Intelligent Decision-Making and Event Management

The distinctive global ID management system within this phase ascertains object identification during communication. The phase involves various activities that require intelligent mechanisms to interact with humans. Therefore, smart algorithms are essential for efficient and effective human interaction. These algorithms manage tasks such as setting up and communicating rules, generating requests, initiating sessions, interacting with diverse interfaces, and terminating sessions.

G. Event Processing and Ontological Modeling

Guided by an ontology, intelligent decisions ensure the targeted distribution of events to relevant departments through unicasting. Events are classified as high-level or low-level, influencing their design, generation, and forwarding to corresponding departments or users. High-level events are stored at the departmental level and directly delivered to recipients, while low-level events remain within the system. The subservice event layer generates events and transmits them to the embedded notification component. Before being fed into the architecture, the data undergo preprocessing to remove noise, ensure uniformity, and eliminate anomalies, accelerating processing. The pre-processed data are then divided into chunks for parallel processing, utilizing distributed storage to support this approach. To handle the vast amounts of the generated multimedia data, a robust architecture is employed. HDFS provides distributed storage, ideal for the map-reduce paradigm used for parallel processing.

The referenced capacity and DP (Dynamic Programming) algorithms play specific roles in processing data within a cluster in the context of a smart community. Capacity algorithms are determining the most efficient way to allocate resources, such as bandwidth, storage, or computational power, to various tasks or services within the cluster. By doing so, they ensure that the system operates within its capacity constraints while maximizing performance. DP algorithms systematically solve these subproblems and store their solutions to avoid redundant computations, ultimately combining them to solve the larger problem efficiently.

Predetermined algorithms (like capacity and DP algorithms) process data within the cluster and the results are then sent to relevant service providers for decision-making. These decisions are ultimately communicated back to the multimedia systems within the smart community for implementation. This closed-loop system ensures efficient data analysis and informed decision-making in managing multimedia resources.

The architecture described is a robust, distributed processing system designed to handle large volumes of data. It begins by preprocessing the data to remove noise, ensure uniformity, and eliminate anomalies, which enhances processing efficiency. The pre-processed data are divided into chunks and processed in parallel, utilizing distributed storage systems to support this scalable and efficient approach, making it well-suited for managing the data demands of a smart community.

The MMBD management module handles the organization, collection, and storage of massive datasets, distributing data across various devices to reduce the load on central servers or the cloud. This intelligent system acquires data from diverse devices via the internet, including sensors, cameras, and object-mounted devices, to capture real-time environmental information. The collected data are analyzed to gain insights and make intelligent decisions, supporting a wide range of multimedia systems and services. This process involves data aggregation, where data are grouped based on connected devices, enabling efficient processing of the massive datasets. The multimedia big data processing utilizes Apache Spark for parallel processing of MMBD, distributed across multiple nodes in a cluster. The system leverages YARN for cluster management, providing a central processing unit for data analysis, including training and inferencing. Prior to processing, raw multimedia data undergo quality control to address issues like irrational data combinations, missing values, and out-of-range values. The architecture employs a parallel and distributed processing unit utilizing MapReduce programming, where data are processed in a distributed manner and stored using a distributed file system.

An optimized MapReduce model is introduced for MMBD analytics, efficiently processing massive datasets in parallel and addressing machine failures, performance issues, and communication challenges. YARN, with its dynamic programming capabilities, manages job distribution and resource allocation within the cluster, improving upon the limitations of traditional Hadoop systems.

IV. RESULTS AND DISCUSSION

The proposed MMBD framework was validated using the open-source platform Apache Hadoop version 3.0, with Apache Spark configured for real-time stream processing. The framework was evaluated using authentic data sets [22]. Preprocessing was performed to remove anomalies and noise from the dataset, leading to significant improvements in processing time and throughput. The introduced architecture demonstrated superior processing time compared to existing solutions for smart city planning, attributed to the optimized YARN cluster management and MapReduce implementation. The former also achieved significantly higher throughput compared to the latter, demonstrating the effectiveness of the customized distributed framework. The optimized YARN, with its dynamic job distribution and resource management capabilities, played a crucial role in enhancing the overall throughput. Figures 3 and 4 display the processing time and output comparison with older methods. Figure 5 exhibits the older versus the new optimized cluster management.

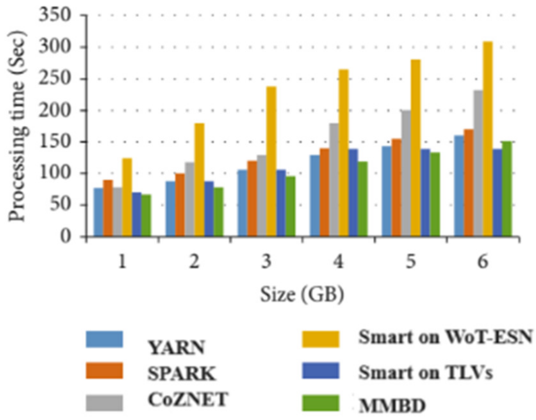

Fig. 3. Processing time comparison with other methods.

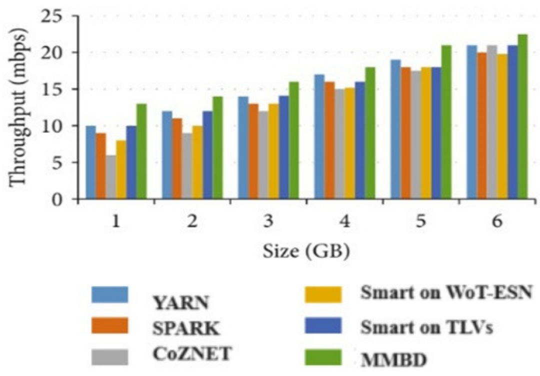

Fig. 4. Throughput comparison with other methods.

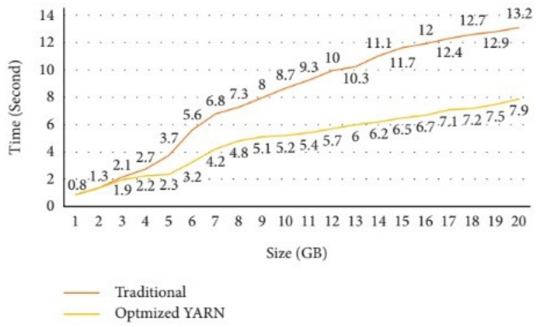

Fig. 5. Older vs. new optimized cluster management.

V. CONCLUSION

The approach proposed in the present study leverages the synergy between Big Data Analytics (BDA) and the Internet of Things (IoT) to create a dynamic and intelligent traffic management system for smart cities. By harnessing real-time traffic data and advanced algorithms, the system effectively optimizes traffic flow, reducing congestion, enhancing mobility, improving road safety, and promoting a more sustainable urban environment. Its adaptive learning capabilities ensure continuous improvement, adapting to evolving traffic patterns and maximizing its efficiency in managing the complexities of urban mobility.

REFERENCES

[1] H. Khayyam, B. Javadi, M. Jalili, and R. N. Jazar, "Artificial Intelligence and Internet of Things for Autonomous Vehicles," in *Nonlinear Approaches in Engineering Applications: Automotive Applications of Engineering Problems*, R. N. Jazar and L. Dai, Eds. New York, NY, USA: Springer, 2020, pp. 39–68.

[2] V. K. Quy *et al.*, "IoT-Enabled Smart Agriculture: Architecture, Applications, and Challenges," *Applied Sciences*, vol. 12, no. 7, Jan. 2022, Art. no. 3396, https://doi.org/10.3390/app12073396.

[3] Z. Lv *et al.*, "Managing Big City Information Based on WebVRGIS," *IEEE Access*, vol. 4, pp. 407–415, 2016, https://doi.org/10.1109/ACCESS.2016.2517076.

[4] A. Ndikumana, S. Ullah, K. Thar, N. H. Tran, B. J. Park, and C. S. Hong, "Novel Cooperative and Fully-Distributed Congestion Control Mechanism for Content Centric Networking," *IEEE Access*, vol. 5, pp. 27691–27706, 2017, https://doi.org/10.1109/ACCESS.2017.2778339.

[5] J. Nie, J. Zhang, W. Ding, X. Wan, X. Chen, and B. Ran, "Decentralized Cooperative Lane-Changing Decision-Making for Connected Autonomous Vehicles," *IEEE Access*, vol. 4, pp. 9413–9420, 2016, https://doi.org/10.1109/ACCESS.2017.2649567.

[6] B. Liu, T. Zhang, and W. Hu, "Intelligent Traffic Flow Prediction and Analysis Based on Internet of Things and Big Data," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, 2022, Art. no. 6420799, https://doi.org/10.1155/2022/6420799.

[7] G. Kotapati, M. A. Ali, and R. Vatambeti, "Deep Learning-Enhanced Hybrid Fruit Fly Optimization for Intelligent Traffic Control in Smart Urban Communities," *Mechatronics and Intelligent Transportation Systems*, vol. 2, no. 2, pp. 89–101, Jun. 2023, https://doi.org/10.56578/mits020204.

[8] A. B. F. Khan and P. Ivan, "Integrating Machine Learning and Deep Learning in Smart Cities for Enhanced Traffic Congestion Management: An Empirical Review," *Journal of Urban Development and Management*, vol. 2, no. 4, pp. 211–221, Dec. 2023, https://doi.org/10.56578/judm020404.

[9] M. W. Levin, S. D. Boyles, J. Duthie, and C. M. Pool, "Demand Profiling for Dynamic Traffic Assignment by Integrating Departure Time Choice and Trip Distribution," *Computer-Aided Civil and Infrastructure Engineering*, vol. 31, no. 2, pp. 86–99, 2016, https://doi.org/10.1111/mice.12140.

[10] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial–temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, Feb. 2016, https://doi.org/10.1016/j.neucom.2015.12.013.

[11] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, May 2018, https://doi.org/10.1007/s11280-017-0487-4.

[12] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 922–929, Jul. 2019, https://doi.org/10.1609/aaai.v33i01.3301922.

[13] Y. Zhang, Q. Cheng, Y. Liu, and Z. Liu, "Full-scale spatio-temporal traffic flow estimation for city-wide networks: a transfer learning based approach," *Transportmetrica B: Transport Dynamics*, vol. 11, no. 1, pp. 869–895, Dec. 2023, https://doi.org/10.1080/21680566.2022.2143453.

[14] G. Sun, R. Qi, Y. Liu, and F. Xu, "A dynamic traffic signal scheduling system based on improved greedy algorithm," *PLOS ONE*, vol. 19, no. 3, Mar. 2024, Art. no. e0298417, https://doi.org/10.1371/journal.pone.0298417.

[15] T. Kabudi, I. Pappas, and D. H. Olsen, "AI-enabled adaptive learning systems: A systematic mapping of the literature," *Computers and Education: Artificial Intelligence*, vol. 2, Jan. 2021, Art. no. 100017, https://doi.org/10.1016/j.caeai.2021.100017.

[16] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, Jul. 2015, https://doi.org/10.1016/j.bushor.2015.03.008.

[17] J. Mills, J. Hu, and G. Min, "Communication-Efficient Federated Learning for Wireless Edge Intelligence in IoT," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5986–5994, Jul. 2020, https://doi.org/10.1109/JIOT.2019.2956615.

[18] M. Elassy, M. Al-Hattab, M. Takruri, and S. Badawi, "Intelligent transportation systems for sustainable smart cities," *Transportation Engineering*, vol. 16, Jun. 2024, Art. no. 100252, https://doi.org/10.1016/j.treng.2024.100252.

[19] Y. Gong, L. Zhang, R. Liu, K. Yu, and G. Srivastava, "Nonlinear MIMO for Industrial Internet of Things in Cyber–Physical Systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5533–5541, Aug. 2021, https://doi.org/10.1109/TII.2020.3024631.

[20] S. Uppoor, O. Trullols-Cruces, M. Fiore, and J. M. Barcelo-Ordinas, "Generation and Analysis of a Large-Scale Urban Vehicular Mobility Dataset," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 1061–1075, May 2014, https://doi.org/10.1109/TMC.2013.27.

[21] M. Khan, B. N. Silva, and K. Han, "Internet of Things Based Energy Aware Smart Home Control System," *IEEE Access*, vol. 4, pp. 7556–7566, 2016, https://doi.org/10.1109/ACCESS.2016.2621752.

[22] "Tabular Playground Series - Mar 2022." [Online]. Available: https://kaggle.com/competitions/tabular-playground-series-mar-2022.