IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 11, November 2024

Cloud-Based Big Data Analytics Platform

[1] Ameet Kumar Mishra, [2] K Srinivas, [3] Harsh Verma, [4] Simran
[1] [2] [3] CSE Department, Chandigarh University, Mohali, India
[4] Assistant Professor, CSE Department, Chandigarh University, Mohali, India
Corresponding Author Email: [1] ameetkumar119@gmail.com, [2] kollasrinivas2402@gmail.com, [3] harshverma4466@gmail.com, [4] simranchawla2027@gmail.com

Abstract— ***This research introduces a novel big data analytics platform that utilizes the power of cloud computing for complex data processing tasks. By leveraging scalable cloud infrastructure and advanced analytics tools, the platform aims to uncover hidden patterns and trends, providing valuable business insights. Key features include processing data in real time, advanced analytics capabilities, and seamless integration with popular cloud services. The platform's effectiveness is demonstrated through a series of experiments, highlighting its ability to manage large-scale datasets and deliver accurate results.***

Keywords— ***Big Data, Cloud Computing, Apache Kafka, Apache Flink, Grafana, PostgreSQL.***

I. INTRODUCTION

In today's digital landscape, the sheer amount of data generated from a multitude of sources—such as social media interactions, IoT devices, and transaction logs—has reached unprecedented levels.The amount, diversity, and velocity of this phenomenon—known as "Big Data"—make standard data processing techniques more inadequate. Additionally, the dimensions of veracity (the trustworthiness of data) and value (the insights that can be extracted) play crucial roles in understanding the complexities of Big Data..

To address these challenges, many organizations are leveraging Cloud Computing as a foundational technology. Cloud platforms offer a flexible and scalable infrastructure that allows for efficient storage and analysis of large datasets without the need for substantial investments in physical hardware. This shift enables organizations to allocate resources dynamically, adapting quickly to their evolving data demands.

The advent of Analytics-as-a-Service (AaaS) can be attributed to the convergence of cloud technology and analysis of big data. Major cloud service providers such as Amazon Web Services, Google Cloud, and Microsoft Azure have developed comprehensive tools that facilitate advanced data analytics across various formats, both structured and unstructured. By utilizing cutting-edge technologies—including distributed storage systems and sophisticated data transformation processes—these platforms enable rapid and insightful analysis.

In a cloud environment, the data analytics workflow typically consists of multiple phases: data collection, storage, processing, and analysis. Data is collected from various sources, and cloud infrastructure supports both real-time and batch processing capabilities. Tools like Apache Spark, Hadoop, and Kafka are pivotal in managing these processes, handling everything from streaming data to large-scale batch operations.

An additional advantage of cloud platforms is their ability to streamline data processing methodologies, transitioning from traditional ETL (Extract, Transform, Load) to ELT (Extract, Load, Transform). This modern approach allows for faster data ingestion and processing directly within the cloud, minimizing the need for extensive data staging and optimizing overall operational efficiency.

In summary, the incorporation of big data analytics and Cloud Computing provides a powerful framework for organizations to derive valuable knowledge from their data. With notable benefits like enhanced scalability, reduced costs, and improved operational flexibility, cloud-based analytics platforms are essential tools in today's data-driven business landscape.

This research paper will introduce a tailored Cloud-Based Big Data Analytics Platform that we are developing, detailing its architectural framework, methodologies, and algorithms. We will also explore its practical applications in solving contemporary data challenges.

II. LITERATURE REVIEW

In the study [1], Farhana Zulkernine et al. propose CLAaaS, a cloud-based Analytics-as-a-Service (AaaS) platform for big data analytics. CLAaaS integrates multi-tenancy, scalable data management, and workflow execution across various analytic tools and data sources. It provides custom interfaces for different user groups and ensures SLAs, promoting web collaboration. Security and data privacy concerns are addressed through private cloud deployment and distributed workflow execution. CLAaaS aims to leverage cloud infrastructure for analytics, facilitating the growing need for scalable, collaborative, and secure big data analytics solutions.

In the study [2], Zaheer Khan et al. discuss the potential of smart cities to enhance urban planning and management through innovative technologies. The gathering, handling, and evaluation of enormous volumes of data produced by many sources, including as sensors, cellphones, and residents, is essential to these cities. This data, when integrated with city repositories, enables effective decision-making for urban governance. The paper emphasizes the need for a uniform, integrated information model in a cloud environment to develop diverse services for city applications. Cloud computing is identified as a crucial solution for managing and analyzing smart city big data, though new tools are required for efficient processing. Future research will focus on developing a prototype to address technical challenges and propose practical solutions.

M. Mayilvaganan and M. Sabitha offer an architecture based on the cloud with vast data processing for smart grids in the study[3]. The system can forecast energy output and demand by analyzing past data and customer behavior using cloud computing paradigms like IaaS, PaaS, and SaaS. The architecture employs the Hadoop Distributed File System (HDFS) for data storage and Cassandra for database management. The proposed system uses a distributed environment to manage renewable and non-renewable energy sources, balancing supply and demand efficiently. Future work will focus on implementing the proposed architecture to improve power production and distribution in smart grids.

In the article [4], Siqi Wu et al. provide SMASH, a big data analysis framework designed to handle large-scale data, especially data pertaining to traffic on roadways. Components such as GeoServer, GeoMesa, Accumulo, Spark, and Hadoop are used in the construction of SMASH. In contrast to Hadoop MapReduce, Spark allows for quicker data processing, while Accumulo and GeoMesa effectively store spatiotemporal data. GeoServer offers customizable map overlays for data visualization. If organizations like VicRoads cooperate, future work will concentrate on increasing cluster size for real-time data aggregation, improving visualization, and gaining access to real-time SCATS traffic data.

In the study[5], Yanish Pradhananga et al. introduce CBA, a platform that utilizes cloud for processing huge amounts of data, focusing on linear regression and time series analysis. The system's distributed framework allows for efficient handling of large-scale data through memory scaling and parallel processing. Combining open-source R and Apache Hadoop, CBA offers scalability, accessibility, and high performance. Future improvements include integrating additional analytics functionalities (e.g., k-means, correlation), reducing the number of nodes for cost-efficiency, and enhancing compatibility with business intelligence tools like Oracle or SAP. CBA can also be deployed on private cloud environments and integrated with HBase for further expansion.

Zhigao Zheng et al. [6] address the issues provided by the increasing development of data from services like cloud computing, IoT, and services according to geographic area by putting forth an agile processing (RTDP) architecture in their research that makes use of cloud technology. The architecture consists of four layers and employs a hierarchical computing model to enhance real-time data processing performance. A multi-level storage framework and an LMA-based application deployment method are described to handle the RTDP system's varied and real-time needs. The system leverages various technologies, including DSMS, CEP, batch-based MapReduce, FPGA, GPU, CPU, and ASIC, for data processing at collection terminals. Data is structured, uploaded to the cloud, and processed using MapReduce within the cloud's powerful computing architecture. The study also presents a general framework and computational methods for future RTDP system designs, positioning it as a foundational approach for real-time big data management.

In the study[7], Amitkumar Manekar and Dr. G. Pradeepini propose a distributed heterogeneous system that is yet to be fully developed. The research highlights the challenges in creating such an application that leverages cloud as a service provider for data storage and analytics from diverse data sources. The study suggests that Hadoop MapReduce is likely to play a pivotal role in addressing these challenges, especially with the integration of advanced cloud features. However, overall data security remains a significant concern when transforming data into the cloud for analytics. The authors emphasize that with advancements in migration algorithms and the adoption of open-source techniques, next-generation computing is set to transition seamlessly to cloud-based infrastructures. Future work will focus on enhancing security measures and optimizing cloud features to enable efficient analytics in distributed environments.

Zaheer Khan and colleagues present a tool that uses cloud technology to process huge amounts of information in smart cities in their study [8]. A prototype for analyzing Bristol Open Data and finding connections between indicators of the urban environment was created using Spark and Hadoop. The service focused on crime, safety, economy, and employment data, with experiments comparing the performance of Hadoop and Spark. The study assesses trends over several years, highlighting both positive and negative outcomes. Future work will focus on scaling the system for broader applications in smart cities.

D. P. Acharjya and Kauser Ahmed P. examine a range of large data analysis research concerns, difficulties, and techniques in their study[9]. The writer's point out that distinct big data systems are designed for particular purposes; some are better suited for batch processing, while others are better suited for real-time analytics. Every platform has features that are specific to it. The analysis sheds light on these platforms' advantages and disadvantages when it comes to meeting certain analytical requirements.

In This paper[10], T. Sri Harsha examines how cloud computing integrates with processing massive data to tackle the growing challenges of data management, focusing on the benefits of scalability, resource sharing, and cost-efficiency. The research highlights platforms like Hadoop for distributed data processing and key cloud features such as virtualization and dynamic resource provisioning. It also addresses big data challenges, categorized under volume, variety, and velocity, and explores different cloud models. The study comes to the conclusion that cloud computing will be fundamental for massive data management in the future and is necessary for processing enormous datasets in an effective manner.

In their study[11], E. Goldin et al.present a cloud computing architecture intended for process control sector big data analytics, addressing the challenges posed by the vast data generated by modern industrial processes. The system integrates real-time data collection from sensors and wireless technologies with cloud storage, facilitating both batch and near-real-time data analysis. Utilizing a Hadoop cluster and Apache Spark for distributed storage and high-speed processing, the architecture enables machine learning algorithms to optimize industrial processes. By leveraging cloud computing, it reduces costs and complexity, making advanced control systems accessible to smaller plants and accommodating the diverse nature of process data. A case study on a walking beam furnace illustrates its capability for dynamic process optimization, marking a significant step toward integrating big data analytics into process control and supporting the transition to Industry 4.0.

Anish Jindal et al. describe in their study[12] an uncertain rule-based big data analytics architecture designed to manage the diverse and significant volumes of healthcare data generated by medical devices and remote body sensors effectively, offering Healthcare-as-a-Service (HaaS) in a cloud computing environment. The architecture uses an ambiguous rule-based classifier to extract and classify medical data according to symptoms reported by medical experts. It has a multi-layered design for data collection, storage, and processing on the cloud. The solution incorporates initial cluster formation and retrieval techniques based on a modified Expectation-Maximization (EM) algorithm, enabling the processing of uncertain health parameters. By leveraging cloud computing, the system facilitates real-time access to patient data, aiding in disease prediction and supporting remote diagnostics for healthcare providers. Analyzed using a range of criteria, including false positive ratio, computation cost, response time, and classification accuracy, the findings show that the architecture consistently provides real-time healthcare data analytics, positioning it as a promising solution for future HaaS platforms.

Rajeswari et al. present a smart agricultural model in their study[13] that addresses issues in agriculture by integrating mobile technologies, cloud-based large-scale data analytics, and IoT. This concept boosts agricultural production by leveraging IoT devices to acquire data in real time, which is then saved in a cloud database. This data is analyzed utilizing big data analysis methods, with an emphasis on important variables including crop growth, market demands, and fertilizer requirements. The insights derived from the analysis are communicated to farmers through mobile applications, enabling them to make informed decisions. By leveraging predictive analytics, data mining, and cloud technologies, the model aims to improve the efficiency of agricultural practices, increase crop production, lower costs, and optimize resource management, thus providing a comprehensive framework for precision agriculture.

Manoj Muniswamaiah et al. examine the advantages and disadvantages of big data applications in cloud computing settings in their study[14]. They emphasize the need for advanced processing techniques due to the rapid increase in data generation from smartphones, sensors, and social media. To address these challenges, the authors propose a scalable architecture that manages both structured and unstructured data by leveraging cloud computing's capabilities, including parallel processing, scalability, virtualization, and data security. They employ tools that are based on cloud like IaaS, PaaS, and SaaS to enhance data processing flexibility and efficiency.Additionally, the authors identify key challenges, including data transmission, security, privacy, and scalability, and provide design principles to optimize big data architecture, positioning this framework as a foundation for future big data processing and cloud management systems.

In their study[15], Ying Liu et al. propose a structure which works on cloud and deals with huge amounts of data aimed at design innovation for SMEs that is motivated by customer insights The framework addresses the challenges SMEs face in utilizing large volumes of online customer data for product design and strategic decision-making. It collects customer reviews, conducts sentiment analysis, and provides insights into customer preferences and product features. The architecture integrates data acquisition, sentiment analysis, and data visualization, enabling SMEs to access advanced analytics tools without substantial IT infrastructure, thus reducing implementation costs and complexity. The study underscores the scalability and flexibility of cloud computing, making big data analytics accessible and affordable for SMEs, ultimately empowering them to improve product development and customer satisfaction while remaining competitive in the market.

Berisha et al. provided a summary of use of Cloud technology when dealing with analysis of huge data in their study[16], highlighting the significance of these technologies in handling massive amounts of heterogeneous, rapidly changing data. They highlight the limitations of traditional processing tools in handling the massive amounts of daily-generated data and demonstrate how cloud computing provides the infrastructure needed for large data processing, storage, and analysis that is both economical and effective. The paper discusses the value of BDA across sectors such as healthcare, education, and business, illustrating how organizations utilize big data for decision-making and process optimization. Using Google’s BigQuery as a case study, the authors showcase its serverless architecture, scalable analysis, and machine learning capabilities while conducting experiments on dataset sizes to evaluate its

IFERP Explore Your Research Journey...
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 11, November 2024

performance with petabytes of data. Furthermore, they advocate for transitioning from the traditional ETL to ELT paradigm, which enhances scalability and flexibility by processing data after its upload to the cloud. Overall, the study positions cloud-based BDA as crucial for future data-driven systems and decision-making.

In their study[17], Rishabh Rajesh Shanbhag et al. examine how process control can be improved across a range of industrial sectors using cloud-based big data analytics, including oil and gas, chemical manufacturing, pharmaceuticals, and food processing. They emphasize how these technologies can improve production efficiency and decision-making by integrating data from multiple sources. The proposed architecture leverages cloud-based storage and advanced analytics through platforms like AWS and Azure to provide real-time insights. Their findings underscore the potential of cloud-based big data analytics to optimize operational performance while also addressing challenges related to data security and system integration.

In the study[18], Amanpreet Kaur Sandhu tackles the issues of handling massive amounts of data generated from various sources, such as social media and Internet of Things devices, and highlights the crucial function of cloud computing in offering scalable solutions for data storage, processing, and analysis. Important cloud services such as Microsoft Azure, Google Cloud, and Amazon Web Services (AWS) are highlighted in the research, which also compares several cloud-based big data frameworks and looks at topics like distributed storage, data security, and data visualization. The paper explores the connection between cloud computing and big data, focusing on how well large data can be managed in cloud environments). It points out important issues including data protection, cleansing, and processing, as well as the difficulties in displaying unstructured data. In the end, it makes the case for the use of cloud computing as an affordable and expandable way to handle the demands of large-scale analytics.

In the study[19], Zaineb Naamane reviews the literature in a methodical manner on big data analysis using cloud, looking at the advantages and disadvantages of integrating cloud computing with big data. The study highlights the ways in which cloud infrastructure makes big data processing, analysis, and storage easier. It also lists benefits including resilience, scalability, and cost savings. It also highlights important obstacles, such as network dependence, security issues, heterogeneity in data, and trouble managing and displaying big information. The article places integration of these technologies as critical for organizations in the digital era, highlighting the need for further research to overcome these difficulties and improve big data analysis using cloud based solutions.

III. ARCHITECTURE

In this architecture, we are creating a cloud-based big data analytics platform that leverages modern technologies for real-time data ingestion, storage, processing, and visualization.

The proposed architecture consists of four key components:

- Data Generation Layer (Apache Kafka):
  - Purpose: Real-time data production and streaming.
  - Components:
    - Apache Kafka: Configured as a distributed messaging system.
    - Data Producers: Generate and send real-time data streams to Kafka topics.
    - Data Schema: Formats like JSON, Avro are used to define the structure of the data.
  - Role: Kafka topics receive data generated by producers for further processing. (Fig. 1)
- Data Ingestion and Processing Layer (Apache Flink):
  - Purpose: Stream processing and real-time transformations.
  - Components:
    - Apache Flink: A distributed stream processing framework.
    - Flink Jobs: Consume data from Kafka topics.
    - Data Transformations: Apply operations like filtering, grouping, windowing, and calculations.
    - Output: Processed data is sent to PostgreSQL for storage and further analysis.
  - Role: This layer performs the core computations, filtering, and processing of incoming data streams.(Fig. 2)
- Data Storage Layer (PostgreSQL):
  - Purpose: Storing processed data for querying and analysis.
  - Components:
    - PostgreSQL Database: Configured to handle high write throughput.
    - Database Tables and Indexes: Created for efficient data storage and optimized retrieval.
  - Role: Stores the processed data coming from Flink and ensures it is accessible for analytics.(Fig. 3)
- Visualization Layer (Grafana):
  - Purpose: Data visualization and dashboard creation.
  - Components:
    - Grafana: Visualization platform connected to the PostgreSQL database.
    - Dashboards: Custom visualizations like charts, graphs, and tables are created based on the data.
  - Role: This layer allows users to explore data insights through interactive dashboards.(Fig. 4)

The proposed architecture for the cloud-based big data analytics platform effectively integrates advanced technologies to facilitate real-time data ingestion, processing,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024

storage, and visualization. By utilizing Apache Kafka for data generation and streaming, Apache Flink for processing and transformations, PostgreSQL for efficient data storage, and Grafana for interactive visualization, the architecture ensures scalable and efficient analytics capabilities. To handle and evaluate enormous volumes of data, each layer is essential for helping firms get insightful knowledge and improve their decision-making procedures. This all-encompassing strategy presents the platform as a strong answer to current data problems across a range of industries.

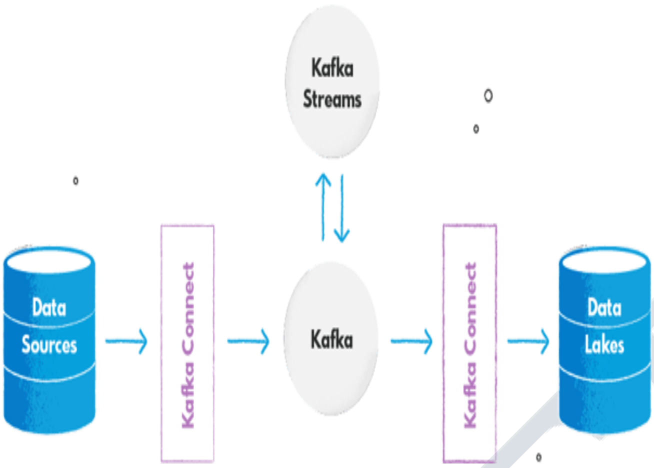

Figure I. Data Generation Layer

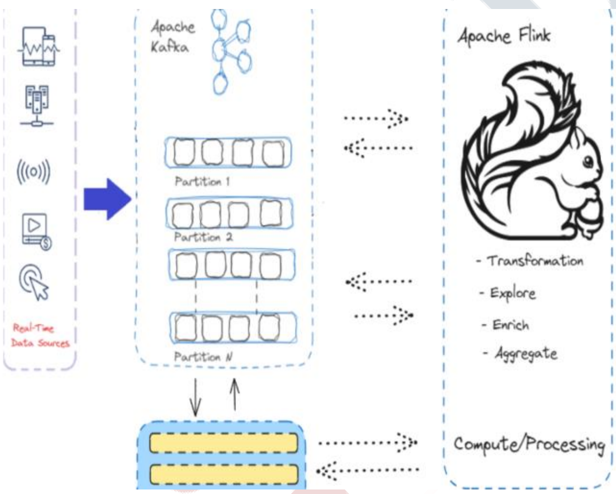

Figure II. Data Ingestion and processing layer

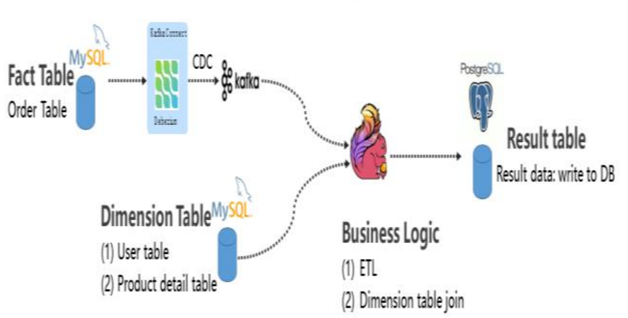

Figure III. Data Storage Layer

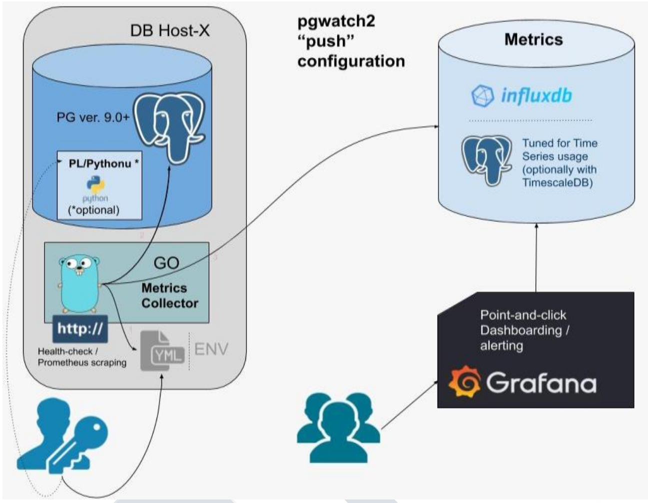

Figure IV. Visualization Layer

IV. ALGORITHM

A cloud-hosted big data analytics platform employs the proposed algorithm, which oversees immediate data processing. It begins with Apache Kafka for streaming data ingestion, followed by Apache Flink for real-time processing, including filtering and aggregation. The processed data is stored in a PostgreSQL database for efficient querying. Finally, Grafana is used to visualize the data through interactive dashboards. This system ensures scalable, real-time data processing and visualization, making it ideal for handling large datasets and delivering actionable insights.

- **Algorithm**:

1. Data Generation:
   - A Python script generates random city data and serializes it in JSON format for Kafka.
2. Data Ingestion:
   - A Kafka producer sends the generated data to a Kafka topic, optimized for performance.
3. Data Batching:
   - Kafka collects and batches incoming data, configured with specific batch size and retention policies.
4. Data Processing:
   - Flink consumes and processes the data, performing aggregations and storing results in PostgreSQL.
5. Data Storage:
   - Processed data is stored in PostgreSQL with efficient indexing for querying.
6. Data Visualization:
   - Grafana connects to PostgreSQL and creates interactive dashboards for visualizing the stored data.

V. RESULTS

We developed a comprehensive big data analysis platform specifically designed to process and analyze meteorological data utilizing cloud-based technologies. The platform begins

IFERP Explore Your Research Journey...

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 11, November 2024

by leveraging a Python script that generates random weather-related data, including unique identifiers (id), city names, and corresponding average temperatures. This data is then ingested by Apache Kafka, a distributed event streaming platform, which efficiently batches the incoming information for further processing.

Once Kafka completes its batching operation, the data is handed off to Apache Flink, a powerful stream processing framework that processes and analyzes large datasets in real time. Flink processes the meteorological data by filtering, aggregating, or performing other data transformation operations. Afterward, the processed data is stored in a PostgreSQL database, which serves as the structured data storage backend. To make the stored data more meaningful, we integrated Grafana, a multi-platform open-source analytics and interactive visualization tool. Grafana provides real-time dashboards and insights into weather patterns, trends, and anomalies based on the processed data. This visualization layer is crucial in allowing users to explore and understand weather data in an intuitive manner, making the platform highly practical for real-time analysis.

The following figures demonstrate the output of the generated data from the Python script and how it is visually represented through the Grafana dashboard, showcasing the flow of data from ingestion to visualization.

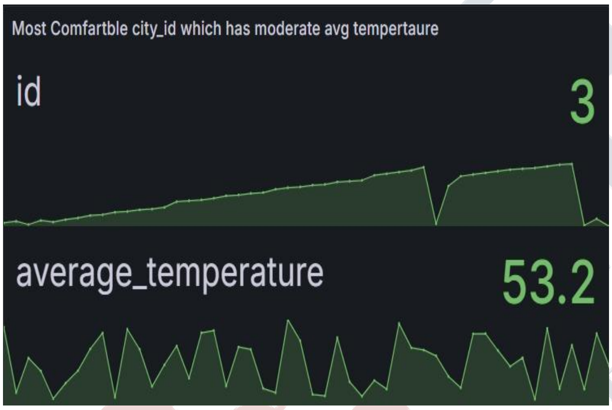

Figure I

This figure tells about the city which has best moderate temperature

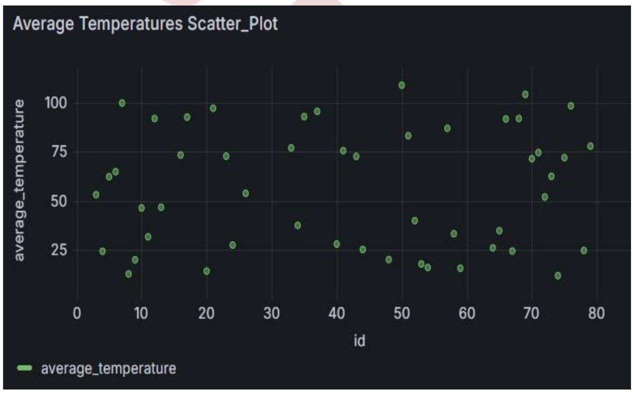

Figure II

This is Scatter plot of cities according to their average temperatures on x-axis = id(id of city) on y-axis(average temperatures)

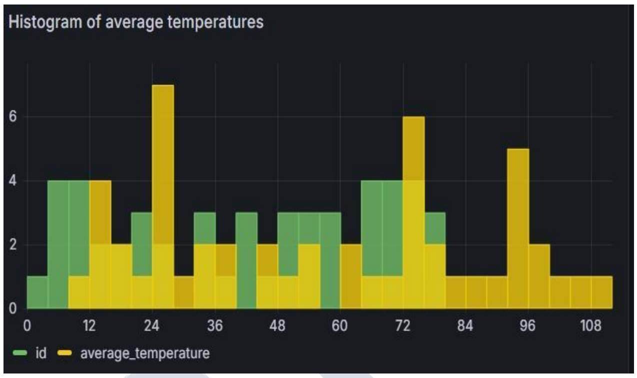

Figure III

VI. CONCLUSION AND FUTURE WORK

In this study, we successfully created a cloud-based big data analysis platform by utilizing scalable architecture to process data in real-time. Using **Python** for data generation, **Apache Kafka** for ingestion and batching, **Apache Flink** for processing, **PostgreSQL** for storage, and **Grafana** for visualization, we demonstrated the platform's capability to handle large, continuous data streams with efficient processing and insightful visualization. The system provides a robust solution for real-time analytics in meteorological data or similar domains.

We intend to improve the platform in our upcoming efforts by incorporating advanced machine learning models for predictive analytics. Additionally, extending support to handle more complex datasets, improving fault tolerance, and integrating more diverse data sources will further increase the platform's robustness and adaptability.

REFERENCES

[1] F. Zulkernine, P. Martin, Y. Zou, M. Bauer, F. Gwadry-Sridhar and A. Aboulnaga, "Towards Cloud-Based Analytics-as-a-Service (CLAaaS) for Big Data Analytics in the Cloud," 2013 IEEE International Congress on Big Data, Santa Clara, CA, USA, 2013, pp. 62-69, doi: 10.1109/ BigData.Congress.2013.18

[2] Z. Khan, A. Anjum and S. L. Kiani, "Cloud Based Big Data Analytics for Smart Future Cities," 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, Dresden, Germany, 2013, pp. 381-386, doi: 10.1109/ UCC.2013.77.

[3] M. Mayilvaganan and M. Sabitha, "A cloud-based architecture for Big-Data analytics in smart grid: A proposal," 2013 IEEE International Conference on Computational Intelligence and Computing Research, Enathi, India, 2013, pp. 1-4, doi: 10.1109/ICCIC.2013.6724168.

[4] R. O. Sinnott, L. Morandini and S. Wu, "SMASH: A Cloud-Based Architecture for Big Data Processing and Visualization of Traffic Data," 2015 IEEE International Conference on Data Science and Data Intensive Systems,

Sydney, NSW, Australia, 2015, pp. 53-60, doi: 10.1109/ DSDIS.2015.35.

[5] Y. Pradhananga, S. Karande and C. Karande, "CBA: Cloud-Based Bigdata Analytics," 2015 International Conference on Computing Communication Control and Automation, Pune, India, 2015, pp. 47-51, doi: 10.1109/ ICCUBEA.2015.18.

[6] Zheng, Zhigao, et al. "Real-time big data processing framework: challenges and solutions." Applied Mathematics & Information Sciences 9.6 (2015): 3169

[7] Khan, Z., Anjum, A., Soomro, K. et al. Towards cloud based big data analytics for smart future cities. J Cloud Comp 4, 2 (2015). https://doi.org/10.1186/s13677-015-0026-8

[8] A. K. Manekar and G. Pradeepini, "Cloud Based Big Data Analytics a Review," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 2015, pp. 785-788, doi: 10.1109/ CICN.2015.160.

[9] Acharjya, Debi Prasanna, and Kauser Ahmed. "A survey on big data analytics: challenges, open research issues and tools." International Journal of Advanced Computer Science and Applications 7.2 (2016): 511-518.

[10] Vashishth, Akshita & Joshi, Sanjay. (2024). The role of big data in shaping the agribusiness industry. International Journal of Research in Agronomy. 7. 116-121. 10.33545/2618060X. 2024.v7.i7Sb.1059.

[11] E. Goldin, D. Feldman, G. Georgoulas, M. Castano and G. Nikolakopoulos, "Cloud computing for big data analytics in the Process Control Industry," 2017 25th Mediterranean Conference on Control and Automation (MED), Valletta, Malta, 2017, pp. 1373-1378, doi: 10.1109/MED.2017. 7984310.

[12] A. Jindal, A. Dua, N. Kumar, A. K. Das, A. V. Vasilakos and J. J. P. C. Rodrigues, "Providing Healthcare-as-a-Service Using Fuzzy Rule Based Big Data Analytics in Cloud Computing," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1605-1618, Sept. 2018, doi: 10.1109/JBHI.2018.2799198.

[13] Garg, Disha & Alam, Mansaf. (2023). Smart agriculture: a literature review. Journal of Management Analytics. 10. 1-57. 10.1080/23270012.2023.2207184.

[14] Muniswamaiah, M., Agerwala, T., & Tappert, C. C. (2019). Big data in cloud computing review and opportunities. International Journal of Computer Science and Information Technology, 11(4), 43-57. https://doi.org/10.5121/ijcsit. 2019.11404

[15] Ying Liu, Anthony Soroka, Liangxiu Han, Jin Jian, Min Tang, Cloud-based big data analytics for customer insight-driven design innovation in SMEs, International Journal of Information Management, Volume 51, 2020, 102034, ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2019.11.002.

[16] Berisha, B., Mëziu, E. & Shabani, I. Big data analytics in Cloud computing: an overview. J Cloud Comp 11, 24 (2022). https://doi.org/10.1186/s13677-022-00301-w

[17] Rishabh Rajesh Shanbhag et al. 2022. Case Studies and Best Practices in Cloud-Based Big Data Analytics for Process Control. International Journal for Research Publication and Seminar. 13, 5 (Nov. 2022), 292–311. DOI:https://doi.org/10. 36676/jrps.v13.i5.1462.

[18] A. K. Sandhu, "Big data with cloud computing: Discussions and challenges," in Big Data Mining and Analytics, vol. 5, no. 1, pp. 32-40, March 2022, doi: 10.26599/BDMA.2021. 9020016.

[19] Naamane, Zaineb. "A systematic literature review: benefits and challenges of cloud-based big data analytics." Issues in Information Systems 24.1 (2023).