KAFKA TOPIC– THE DATA STORER

Dayyala Aparna[1], V.Navya[2], S.Sindhuja[3], P.Sindhuja[4], T.Abhinav[5], SK.Sameer[6]

[1]Assistant Professor, Department of CSE, Balaji Institute of Technology & Science, Laknepally, Warangal, India

[2,3,4,5,6] BTech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

Abstract The new technology of storing the data is replacing the current database system .The topic in kafka is the future of database . It reduces the space complexity of storage of data, it stores data as a series of events while other database's co-locate data based on the keys and indexes to facilitate fast lookup. Kafka is a messaging system, allowing for messages to be posted by publishers and read by subscribers. Kafka is also a database that provides ACID properties. However, it works differently than other databases.

In Kafka, topics are the categories used to organize messages.Each topic has a name that is unique across the entire kafka cluster. Kafka topics are multi-subscriber. This means that a topic can have zero, one, or multiple consumers subscribing to that topic and the data written to it.

Keywords: Data Storage, Kafka Topic,Storage domains, Unlimited data storing, Ensuring data, Producer, Consumer.

I. INTRODUCTION

Kafka topics are the categories used to organize messages. Each topic has a name that is unique across the entire Kafka cluster.

Messages are sent to and read from specific topics. In other words, producers write data to topics, and consumers read data from topics.

Kafka topics are multi-subscriber. This means that a topic can have zero, one, or multiple consumers subscribing to that topic and the data written to it.

In Kafka, topics are partitioned and replicated across brokers throughout the implementation. Brokers refer to each of the nodes in a Kafka cluster. The partitions are important because they enable parallelization of topics, enabling high message throughput.[1-29]

II. Literature Survey

1. In the realm of Apache Kafka, a "topic" serves as a logical category or stream of data, acting as a container for messages published by producers and consumed by consumers, and can be partitioned for scalability and parallel processing.

Doi:10.48047/ijiee.2025.15.5.53

2. Kafka Topic Storage-Based Approaches

With the rise of limited data storage capacity problem, the Statement lead to the development of the Kafka topic, that stores the data in the form of messages. The development meant to be the approach of data streaming to manage huge data.

Key Research Contributions in Kafka Topic messages storing:

1. C. N. Nguyen, J.-S. Kim, and S. Hwang, ‘‘KOHA: Building aKafka-based distributed queue system on the fly in a Hadoop cluster,’’ inProc. IEEE 1st Int. Workshops Found. Appl. Self Syst. (FAS∗W), Sep. 2016,pp. 48–53
2. H. Wu, Z. Shang, and K. Wolter, ‘‘TRAK: A testing tool for studying thereliability of data delivery in Apache Kafka,’’ in Proc. IEEE Int. Symp.Softw. Rel. Eng. Workshops (ISSREW), Oct. 2019, pp. 394–397.
3. T. P. Raptis, A. Passarella, and M. Conti, ‘‘Distributed data access inindustrial edge networks,’’ IEEE J. Sel. Areas Commun., vol. 38, no. 5,pp. 915–927, May 2020.
4. A. Carnero, C. Martín, D. R. Torres, D. Garrido, M. Díaz, and B. Rubio,‘‘Managing and deploying distributed and deep neural models throughKafka-ML in the cloud-to-things continuum,’’ IEEE Access, vol. 9,pp. 125478–125495, 2021.
5. K. C. Okafor, M. C. Ndinechi, and S. Misra, ‘‘Cyber-physical networkarchitecture for data stream provisioning in complex ecosystems,’’ Trans.Emerg. Telecommun. Technol., vol. 33, no. 4, p. e4407, 2022.

Notable Works in Kafka Topicdata storage

1. G. Hesse and M. Lorenz, ‘‘Conceptual survey on data stream processingsystems,’’ in Proc. IEEE 21st Int. Conf. Parallel Distrib. Syst. (ICPADS),Dec. 2015, pp. 797–802.
2. H. Zhang, L. Fang, K. Jiang, W. Zhang, M. Li, and L. Zhou, ‘‘Secure dooron cloud: A secure data transmission scheme to protect Kafka’s data,’’ inProc. IEEE 26th Int. Conf. Parallel Distrib. Syst. (ICPADS), Dec. 2020,pp. 406–413.
3. G. Hesse, C. Matthies, and M. Uflacker, ‘‘How fast can we insert?An empirical performance evaluation of Apache Kafka,’’ in Proc.IEEE 26th Int. Conf. Parallel Distrib. Syst. (ICPADS), Dec. 2020,pp. 641–648.
4. C. Giblin, S. Rooney, P. Vetsch, and A. Preston, ‘‘Securing Kafkawith encryption-at-rest,’’ in Proc. IEEE Int. Conf. Big Data (Big Data),Dec. 2021, pp. 5378–5387.
5. D. Landau, X. Andrade, and J. G. Barbosa, ‘‘Kafka consumer groupautoscaler,’’ 2022, arXiv:2206.11170.

Doi:10.48047/ijiee.2025.15.5.53

III. Existing System

The current application for data storingis often praised for it’s strong privacy focus and generous free storage, offering 20GB of storage with a free account and end-to-end encryption.

However, some users find the paid plans expensive and the web interface can be slow, especially when uploading many files. It allows users to share files with encrypted links, ensuring that only those with the key can access the data.

2. Workflow of Existing Systems

Upload

1. File selection: User selects files to upload from their device.
2. Encryption: Mega encrypts the files using the user's encryption key.
3. Upload: The encrypted files are uploaded to Mega's servers.
4. Storage: The uploaded files are stored on Mega's servers.

Download

1. File selection: User selects files to download from their Mega account.
2. Decryption: Mega decrypts the files using the user's encryption key.
3. Download: The decrypted files are downloaded to the user's device.

Sharing

1. File selection: User selects files to share from their Mega account.
2. Link generation: Mega generates a link to the selected files.
3. Link sharing: The user shares the link with others.
4. Access control: Mega controls access to the shared files based on user permissions.

3. Technologies Used in Existing Systems

Most existing systems are implemented using:

1. Distributed Storage: Existing System uses a distributed storage system to store files across multiple servers.
2. Redundancy: Existing System uses redundancy to ensure data availability and durability.
3. Erasure Coding: Existing System uses erasure coding to reconstruct data in case of server failures.

4. Limitations of Existing Systems

Doi:10.48047/ijiee.2025.15.5.53

1. Storage costs: Application storage costs can be high, especially for large amounts of storage.

2. Bandwidth costs: Applications's bandwidth costs can be high, especially for large amounts ofdata transfer.

3. Premium features: Applications's premium features can be expensive, especially for individuals or small businesses.

IV. Proposed System

The proposed system is a data storage application, that stores the data as messages in a sequence manner. It uses postman for the data producing that is directly mapped to the broker with the topics.

The project adds the following dependencies (groupId and ArtifactId):

org.springframework.boot , spring-boot-starter-web
org.springframework.kafka , spring-kafka
org.springframework.boot , spring-boot-starter-test
org.springframework.kafka , spring-kafka-test

The dependencies of kafka helps us to provide simple and consistent programming model for working with Apache Kafka. It enables developers to easily integrate Kafka into our Spring-based applications.

The most important key feature used from Kafka configuration is Kafka Template. It helps us to create the topic within the broker that automatically generates the broker to generate topic. It is a simple and convenient way to send messages to kafka topics.

The topics from Apache Kafka are used the proposed system, that stores the data in the form of messages. It stores unlimited messages and consumes zero space in the system.
This helps in reducing space complexity.

V. Problem Statement

The primary challenge every organisation facing is the space consumed by the system to store the data. Every data storer is free for limited amount of data and then to have continuation of storing the data the premium have to be taken for any data storage applications. To overcome this , a project have to be developed that allows users to store the data unlimitedly dispite of having any limit or consuming the memory in the system. As Apache Kakfa is used for datastreaming , it can also used for data storing .

Doi:10.48047/ijiee.2025.15.5.53

VI. Methodology

The proposed system follows a structured approach to achieve real-time solution for the data storing that consumes zero space in the system.

It should include the following steps, defining the topic name and it's purpose of the topic and also determining the structure and format is important to store the data. And next create the topic using Kafka configuration that helps in creating the topic using command-line interface, and should include the topic configuration, topic name, partition count.

Now using mapping method is used to map the data producer, here in the project the mapping is given to the postman service that helps in pushing the data. Use the Kafka consumer to consume the data that is available in the corresponding dependencies that are used, it helps to read the data from the topic.

The project provides high-throughput, horizontal scaling, Fault tolerance, multi-protocol support, low latency, encryption, authentication and it is also open source to use.

VII. Future Scope

Here are some potential future scopes of Kafka topics:

Real-time Data Processing

1. Increased adoption: Kafka will continue to be adopted by more companies for real-time data processing.

2. Improved performance: Kafka's performance will continue to improve, enabling faster and more efficient data processing.

3. Real-time analytics: Kafka will be used for real-time analytics, enabling businesses to make data-driven decisions.

Event-Driven Architecture

1. Event-driven architecture: Kafka will be used as a central component of event-driven architectures.

2. Microservices: Kafka will be used to enable communication between microservices.

3. Event sourcing: Kafka will be used for event sourcing, enabling businesses to store and manage events.

Artificial Intelligence and Machine Learning

Doi:10.48047/ijiee.2025.15.5.53

1. AI and ML integration: Kafka will be used to integrate AI and ML into real-time data processing pipelines.

2. Predictive analytics: Kafka will be used for predictive analytics, enabling businesses to make predictions based on real-time data.

3. Anomaly detection: Kafka will be used for anomaly detection, enabling businesses to detect unusual patterns in real-time data.

Security and Governance

1. Security: Kafka will continue to improve its security features, enabling businesses to protect their data.

2. Governance: Kafka will be used for data governance, enabling businesses to manage and regulate their data.

3. Compliance: Kafka will be used to enable compliance with regulatory requirements, such as GDPR and HIPAA.

VIII. Conclusion

In conclusion, Kafka topics are fundamental to organizing and distributing data within a real-time data stream, acting as a log of events, enabling efficient data handling and processing for various applications.

A Kafka topic is essentially a log of events or data streams, allowing for the storage and retrieval of data in a structured and efficient manner.

Topics enable the organization of data by allowing you to create different topics for different types of events or filtered/transformed versions of the same events.

Kafka topics can be partitioned, meaning that data can be divided into multiple segments for parallel processing, and messages can be sequenced by time, ensuring data integrity.

This project demonstrates the capacity of Kafka Topic that stores huge amount of data. This project is meant to allow the users to access the unlimited storage of the data. This consumes zero bytes of space in the system.

Doi:10.48047/ijiee.2025.15.5.53

REFERENCE

1. C. N. Nguyen, J.-S. Kim, and S. Hwang, ‘‘KOHA: Building aKafka-based distributed queue system on the fly in a Hadoop cluster,’’ inProc. IEEE 1st Int. Workshops Found. Appl. Self Syst. (FAS∗W), Sep. 2016,pp. 48–53
2. H. Wu, Z. Shang, and K. Wolter, ‘‘TRAK: A testing tool for studying thereliability of data delivery in Apache Kafka,’’ in Proc. IEEE Int. Symp.Softw. Rel. Eng. Workshops (ISSREW), Oct. 2019, pp. 394–397.
3. T. P. Raptis, A. Passarella, and M. Conti, ‘‘Distributed data access inindustrial edge networks,’’ IEEE J. Sel. Areas Commun., vol. 38, no. 5,pp. 915–927, May 2020.
4. A. Carnero, C. Martín, D. R. Torres, D. Garrido, M. Díaz, and B. Rubio,‘‘Managing and deploying distributed and deep neural models throughKafka-ML in the cloud-to-things continuum,’’ IEEE Access, vol. 9,pp. 125478–125495, 2021.
5. K. C. Okafor, M. C. Ndinechi, and S. Misra, ‘‘Cyber-physical networkarchitecture for data stream provisioning in complex ecosystems,’’ Trans.Emerg. Telecommun. Technol., vol. 33, no. 4, p. e4407, 2022.
6. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima” A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method” Computers and Electrical Engineering , 101 (2022) 107960
7. Ramdas Vankdothu,.Mohd Abdul Hameed “Adaptive features selection and EDNN based brain image recognition on the internet of medical things”, Computers and Electrical Engineering , 103 (2022) 108338.
8. Ramdas Vankdothu,.Mohd Abdul Hameed,Ayesha Ameen,Raheem,Unnisa “ Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network” Computers and Electrical Engineering,102(2022) 108196.
9. Ramdas Vankdothu,.Mohd Abdul Hameed” Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning” Measurement: Sensors Journal,Volume 24, 2022, 100440 .

Doi:10.48047/ijiee.2025.15.5.53

10. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor MRI images identification and classification based on the recurrent convolutional neural network" Measurement: Sensors Journal,Volume 24, 2022, 100412 .

11. Bhukya Madhu, M.Venu Gopala Chari, Ramdas Vankdothu,.Arun Kumar Silivery,Veerender Aerranagula " Intrusion detection models for IOT networks via deep learning approaches " Measurement: Sensors Journal,Volume 25, 2022, 100641

12. Mohd Thousif Ahemad ,Mohd Abdul Hameed, Ramdas Vankdothu" COVID-19 detection and classification for machine learning methods using human genomic data" Measurement: Sensors Journal,Volume 24, 2022, 100537

13. S. Rakesh [a], NagaratnaP. Hegde [b], M. VenuGopalachari [c], D. Jayaram [c], Bhukya Madhu [d], MohdAbdul Hameed [a], Ramdas Vankdothu [e], L.K. Suresh Kumar "Moving object detection using modified GMM based background subtraction" Measurement: Sensors ,Journal,Volume 30, 2023, 100898

14. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data" Journal of Adv Research in Dynamical & Control Systems, Vol. 12, 2020.

15. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network" International Journal of Advanced Science and Technology, Vol. 29, No. 6, (2020), pp. 2875 – 2881

16. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing", Journal Of Critical Reviews,Vol 7, Issue 08, 2020(Scopus Indexed)

17. Ramdas Vankdothu, Dr.Mohd Abdul Hameed "A Security Applicable with Deep Learning Algorithm for Big Data Analysis",Test Engineering & Management Journal,January-February 2020

18. Ramdas Vankdothu, G. Shyama Chandra Prasad " A Study on Privacy Applicable Deep Learning Schemes for Big Data" Complexity International Journal, Volume 23, Issue 2, July-August 2019

Doi:10.48047/ijiee.2025.15.5.53

19. Ramdas Vankdothu, Dr.Mohd Abdul Hameed, Husnah Fatima " Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network" The International journal of analytical and experimental modal analysis, Volume XII, Issue IV, April/2020

20. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things " Journal of Engineering Sciences, Vol 11,Issue 4 , April/ 2020(UGC Care Journal)

21. Ramdas Vankdothu, Dr.Mohd Abdul Hameed " Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics", Complexity International Journal , Volume 24, Issue 01, Jan 2020

22. Ramdas Vankdothu, G. Shyama Chandra Prasad" A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools" International Journal For Innovative Engineering and Management Research,Vol 08 Issue08, Aug 2019.

23. Vankdothu, R., Hameed, M.A. "An Effective Congestion and Interference Secure Routing Protocol for Internet of Things Applications in Wireless Sensor Network " Wireless Personal Communication Journal 140, 143–161 (2025)

24. Vankdothu, R., Bhukya, H. & Bhukya, R.R. "Hybrid TDR-MI Based Wireless Sensor Network for Underground Water Pipeline Leakage Detection and Localization Using Pressure Residuals and Classifiers Wireless Personal Communications 139, 803–823 (2024).

25. Vankdothu, R., Cheng, X. "Energy Efficient TDMA and Secure Based MAC Protocol for WSN Using AQL Coding and ASGWI Clustering". Wireless Personal Communications 136, 2125–2143 (2024)

26. Vankdothu, R., Hameed, M.A., Fatima, H. *et al.* Multicast Scaling in Heterogeneous Wireless Sensor Networks for Security and Time Efficiency. Wireless Personal Communications (2025).

27. Vankdothu, R., Hameed, M.A., Fatima, H. *et al.* Multicast Scaling in Heterogeneous Wireless Sensor Networks for Security and Time Efficiency. Wireless Personal Communications (2025)

Doi:10.48047/ijiee.2025.15.5.53

28. Ramdas Vankdothu, Mohd Abdul Hameed" Brain MRI Images for Tumor Detection using Storage Optimization Technique",Mobile Radio Communications and 5G Networks,Lecture Notes in Networks and Systems,425-437, Springer .

29. Bandi Krishna , Ramdas Vankdothu , Varun Revuri and B. Prashanth" A brain tumor identification using convolution neural network in the deep learning" MATEC Web of Conferences 392, 01131 (2024) ,https://doi.org/10.1051/matecconf/202439201131 ICMED 2024

Bibliography

I am Navya Vattipally from department of Computer Science and Engineering. Currently, pursuing 3rd year at Balaji Institute of Technology and Science. My research is done on "Kafka Topic".

I am Sheelam Sindhuja from department of Computer Science and Engineering. Currently, pursuing 3rd year at Balaji Institute of Technology and Science. My research is done on "Kafka Topic".

I am Puli Sindhuja from department of Computer Science and Engineering. Currently, pursuing 3rd year at Balaji Institute of Technology and Science. My research is done on "Kafka Topic".

Doi:10.48047/ijiee.2025.15.5.53

I am T Abhinav from department of Computer Science and Engineering. Currently, pursuing 3rd year at Balaji Institute of Technology and Science. My research is done on “Kafka Topic”.

I am SK Sameer from department of Computer Science and Engineering. Currently, pursuing 3rd year at Balaji Institute of Technology and Science. My research is done on “Kafka Topic”.

Doi:10.48047/ijiee.2025.15.5.53