

DR. AMBEDKAR INSTITUTE OF TECHNOLOGY

(An Autonomous institute affiliated to Vishvesvaraya Technology University, Belagavi, Accredited by NAAC, UGC with 'A' Grade) Near Jnana Bharathi Campus, Bengaluru – 560056



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(Accredited by NBA)

“Big Data Analytics”

(SUBJECT CODE – 21CST701)

REPORT

On

“Introduction to Apache Kafka”

Submitted in partial fulfilment of award of the degree of

BACHELOR OF ENGINEERING

SEMESTER: 7

ACADEMIC YEAR: 2024-2025

SUBMITTED BY: -

SI NO	NAME	USN
1	Hajaratali S Mogalalli	1DA22CS409
2	Arun G Koravanavar	1DA21CS027
3	Arya N D	1DA21CS028
4	Bharatkumar N Medegar	1DA21CS033
5	Arjun Singh Pundir	1DA21CS026
6	Ayushman Sharan	1DA21CS029
7	Abhay Singh	1DA21CS004
8	Akhil Kumar Tiwari	1DA21CS014

UNDER THE GUIDANCE OF

Dr.Nandini n

Professor Dept. of CSE

Dr. AIT

Bengaluru-56

Dr. AMBEDKAR INSTITUTE OF TECHNOLOGY
MALLATHAHALLI, OUTER RING ROAD, BENGALURU – 560056
2024-25

Dr. AMBEDKAR INSTITUTE OF TECHNOLOGY
Mallathahalli, Bengaluru – 560056
Department of computer science and engineering



Certificate

Certified that the project work entitled “**Introduction to Apache Kafka**” carried out by Hajaratali S. Mogalalli, bearing USN: 1DA22CS409, Arun G. Koravanavar, bearing USN: 1DA21CS027, Arya N. D., bearing USN: 1DA21CS028, Bharatkumar N. Medegar, bearing USN: 1DA21CS033, Arjun Singh Pundir, bearing USN: 1DA21CS026, Ayushman Sharan, bearing USN: 1DA21CS029, Abhay Singh, bearing USN: 1DA21CS004, and Akhil Kumar Tiwari, bearing USN: 1DA21CS014, Bonafide students of Dr. Ambedkar Institute of Technology, Bangalore – 560056, in partial fulfillment for the award of Bachelor of Engineering in **Computer Science and Engineering** of the Visvesvaraya Technological University, Belagavi, during the year 2022–2023. It is certified that all the corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. Group activity report has been approved as it satisfies the academic requirements.

Signature of Guide

Dr.Nandini n
Professor
Dept. of CSE
Dr.AIT

Signature of HOD

Dr.Nandini n
Professor & Head
Dept. of CSE
Dr.AIT

ACKNOWLEDGEMENT

We would like to express our deep sense of gratitude to our institution **Dr. AMBEDKAR INSTITUTE OF TECHNOLOGY**, for having well qualified staff and well-furnished laboratories with necessary equipment.

We express our deepest gratitude and sincere thanks to Assist. **Dr.Nandini N Professor. of CSE** for their valuable guidance during the course of this project. Her continuous suggestion has helped in making this project a successful one.

We are grateful to our Principal **Dr. M.N.Thippeswamy** for providing necessary time and facilities to carry out the project.

We also express our sincere thanks and heart full gratitude to our Head of the Department **Dr.Nandini N** Department of Computer Science, for his continuous support throughout the project.

We take this opportunity to express our deep sense of gratitude to all the staff members of Department of Computer Science Engineering for providing us with all the required support on and ruff classroom.

Last but not least we express our sincere thanks to all of our friends and our parents who have patiently extended all sorts of help for accomplishing this undertaking.

Hajaratali S Mogalalli
Arun G Koravanavar
Arya N D
Bharatkumar N Medegar
Arjun Singh Pundir
Ayushman Sharan
Abhay Singh
Akhil Kumar Tiwari

ABSTRACT

Apache Kafka is a distributed event streaming platform designed for high-throughput, fault-tolerant, and scalable data processing. Initially developed by LinkedIn and later open-sourced by the Apache Software Foundation, Kafka has become a critical infrastructure tool for modern data-driven organizations. Its ability to handle real-time data streams efficiently makes it ideal for applications such as real-time data pipelines, event sourcing, and log aggregation.

This report delves into the core architecture of Kafka, including its fundamental components such as producers, consumers, topics, partitions, brokers, and Zookeeper (or KRaft in newer versions). Kafka's unique features—such as horizontal scalability, durability, and stream processing capabilities—are explored in detail, highlighting its relevance in addressing contemporary challenges in big data and real-time analytics.

Additionally, the report examines Kafka's use cases across industries and compares it to other popular messaging systems like RabbitMQ, ActiveMQ, and AWS Kinesis, emphasizing Kafka's advantages in scalability, throughput, and flexibility. Installation guidelines and a practical example of Kafka's application in real-time analytics provide insights into its operational usage.

In conclusion, the report underscores Kafka's pivotal role in the modern computing landscape, where real-time data processing is essential for decision-making and innovation. The insights provided serve as a comprehensive guide for understanding Kafka's architecture, functionality, and impact.

Dr. AMBEDKAR INSTITUTE OF TECHNOLOGY, BENGALURU-56

**(An Autonomous Institution Affiliated to VTU, Belagavi, accredited by NAAC with Grade "A"
and accredited by NBA)**

**Department of Computer Science and Engineering
“Big Data Analytics”**

Index Page

SI NO	TOPICS
1	Introduction
2	Fundamentals of Apache Kafka
3	Kafka Architecture
4	Use Cases of Apache Kafka
5	Kafka vs Other Messaging Systems
6	Conclusion
7	References

