**CITY FLOW**

**A Scalable Big Data Framework for Real-Time Analytics in Computing 4.0**

**1Priyanshu Mohanty 2Bishnu Prasad Kar 3Mrutunjay Jena 4Pragnya Priyadarshini Upadhyaa**

**Student, Department of Computer Science & Engineering, 5th Semester, B.Tech,**

**GIET Ghangapatana, Bhubaneswar**

**E-mail:** [**priyanshumohanty2611@gmail.com**](mailto:priyanshumohanty2611@gmail.com)**,** [**karbishnu2430@gmail.com**](mailto:karbishnu2430@gmail.com), [**mrutunjayjena2005@gmail.com**](mailto:mrutunjayjena2005@gmail.com), [**upadhyaypragnyapriyadarsini@gmail.com**](mailto:upadhyaypragnyapriyadarsini@gmail.com)

**ABSTRACT**

The era of Computing 4.0 has accelerated the growth of data in volume, velocity, and variety, driving demand for scalable, intelligent architectures that deliver timely insights. Traditional cloud-based systems often struggle with real-time workloads, particularly in data-intensive domains such as smart cities and intelligent transportation. This paper presents **CityFlow**, a scalable big data framework that unifies distributed ingestion, storage, processing, and machine learning. CityFlow is structured into multiple layers: a data ingestion layer using stream processors like Kafka or Kinesis; a distributed storage layer with HDFS or S3; a processing layer powered by Apache Spark clusters; and a feature store integrating ML pipelines for both batch and streaming analytics. To validate the framework, a proof-of-concept was implemented with the **NYC Taxi dataset**, simulating high-volume mobility data. Results show efficient end-to-end analytics, from ingestion and preprocessing to predictive modeling and real-time visualization. CityFlow demonstrates strong potential for Computing 4.0 applications in smart mobility, predictive demand forecasting, and anomaly detection in transportation systems.