International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 28 March - April - 2026

Optimizing Real-Time Big Data Pipelines: A Comparative Analysis of Edge Computing vs. Cloud Architectures

Dr. Sagar S. Kuthe
Assistant Professor, Department of Statistics,
Brijlal Biyani Science College, Amravati, Sant Gadge Baba Amravati University, Amravati (MS).
Corresponding Author – Dr. Sagar S. Kuthe
DOI - 10.5281/zenodo.19396718

Abstract:

The exponential growth of Internet of Things (IoT) devices has created a paradigm shift in data processing, moving from centralized cloud architectures toward decentralized edge computing. This research investigates the performance optimization of real-time big data pipelines by comparing traditional cloud-based processing with an edge-cloud hybrid model. Utilizing a discrete-event simulation methodology, the study analyses 1,000,000 simulated smart city data events to measure two critical metrics: end-to-end latency and network bandwidth consumption. The results demonstrate that the edge-cloud hybrid model reduces average processing latency by 75%, from 180 ms to 45 ms, while simultaneously decreasing bandwidth usage by approximately 77% through localized data filtering. The findings suggest that while edge computing significantly enhances responsiveness for time-sensitive applications—such as autonomous systems and real-time healthcare monitoring—it is most effective when integrated into a hybrid framework that leverages the cloud for long-term storage and heavy computational modelling. This paper concludes that decentralized architectures are essential for scaling the next generation of big data infrastructure, providing a roadmap for future research into localized AI decision-making.

Keywords: Big Data Analytics, Edge Computing, Cloud Computing, IoT, Latency Optimization, Data Pipelines.

Introduction:

The global data volume is projected to exceed 180 zettabytes by 2025, driven by the proliferation of IoT devices. Traditional cloud-based architectures face significant challenges in processing this data due to network latency, bandwidth bottlenecks, and high data transfer costs. Edge computing has emerged as a decentralized alternative that processes data closer to its source, promising real-time responsiveness for critical applications like autonomous vehicles and smart cities.

Research Objectives:

1. To evaluate the latency reduction achieved by edge computing compared to traditional cloud models.
2. To analyse the impact of decentralized processing on network bandwidth consumption.
3. To identify the trade-offs between local processing power and centralized analytical depth.

Methodology:

This study employed a **Python-based simulation** using libraries such as SimPy for discrete-event modelling and Matplotlib for visualization. We simulated two architectural models:

1. **Cloud-Only Model:** Data is ingested at the source and transmitted to a centralized data centre for processing.
2. **Edge-Cloud Hybrid Model:** Initial filtering and anomaly detection occur at the edge, while long-term storage and complex modelling are offloaded to the cloud.

Data Sources: Simulated traffic sensor data comprising 1,000,000 events to mimic a smart city environment.

Results:

Our simulation revealed a drastic reduction in end-to-end latency when moving processing tasks to the edge.

- **Latency Analysis:** The average processing delay dropped from **180 ms (Cloud)** to **45 ms (Edge)**, representing a **75% improvement**.
- **Bandwidth Efficiency:** Edge filtering reduced total data transmission from **220 Mbps to 50 Mbps**.

Discussion:

The results confirm that edge computing is indispensable for time-sensitive applications where milliseconds matter, such as autonomous driving or real-time health monitoring. However, we observed that edge devices are limited by constrained computational resources, meaning they cannot replace the cloud for high-intensity tasks like training large-scale deep learning models. The "sweet spot" for modern big data lies in hybrid architectures that balance local speed with global intelligence.

Conclusion:

Decentralized big data analytics significantly enhances system performance by reducing latency by up to 75% and optimizing bandwidth usage. Future research should focus on Explainable AI (XAI) at the edge to ensure that localized, automated decisions are transparent and ethically sound.

References:

1. Satyanarayanan, M. (2017). The emergence of edge computing. *Computer*, 50(1), 30-39.
2. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637-646.
3. Ramaswamy, Y., et al. (2024). Edge computing and its impact on information technology. *International Journal of Computer Network and Information Security (IJCNIS)*, 16(5), 744-758.