Migration Letters

Volume: 20, No: S4(2023), pp. 209-215

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)
www.migrationletters.com

Stay Ahead of The Curve: A Deep Dive into Performance Analysis of Streaming Data with Big Data Frameworks

Sampath Kini K[1], Karthik Pai B H[2]

Abstract

Organisations are scrambling to use big data frameworks to harness the power of streaming data in the age of real-time data. In this article, the complex dynamics of real-time data processing optimisation are explored. Offering useful insights and best practises, it explores performance analysis across dimensions like scalability, resource optimisation, latency reduction, fault tolerance, and benchmarking. This research enables organisations to navigate the changing data landscape, make wise technology decisions, and maintain an advantage in the race for real-time insights by bridging the gap between the exponential growth of streaming data and the need for effective analytics.

Keywords: *Streaming data, big data, real time analysis, Resource.*

INTRODUCTION

Continuous data that can be produced and processed in real time is referred to as streaming data. Several sources, including Internet to Things devices, social networking sites, sensors, logs, and more, can produce this data. Data is processed in real-time using the streaming data framework, big data frameworks are made to manage large volumes of data. As it enables individuals to process data that comes in, it is a potent tool for assisting in analysing massive amounts of data. Streaming enables one to interpret data in real-time or very close to real-time, which makes it suitable for a variety of applications including real-time analysis, fraud detection, tracking, and more.

STATISTICAL DATA ANALYSIS

The global big data market is huge and it is growing day by day, a lot use of data in every field needs to be managed. Organisations are managing large quantities of data and for doing this they are taking the help of the frameworks. As per the view of [1], the use of big data is valued at $169 billion in the year 2018. The big data analytics market is anticipated to grow from about $241 billion in 202 to over $655 billion by 2029 [2]. The use of streaming data has seen a rising popularity as the use of social media and online gaming has increased lately.

[1] NMAM Institute of Technology (Nitte Deemed to be University)/CSE Department,Nitte, Karkala, India, Email: sampath@nitte.edu.in

[2] NMAM Institute of Technology (Nitte Deemed to be University)/ISE Department,Nitte, Karkala, India, Email: karthikpai@nitte.edu.in

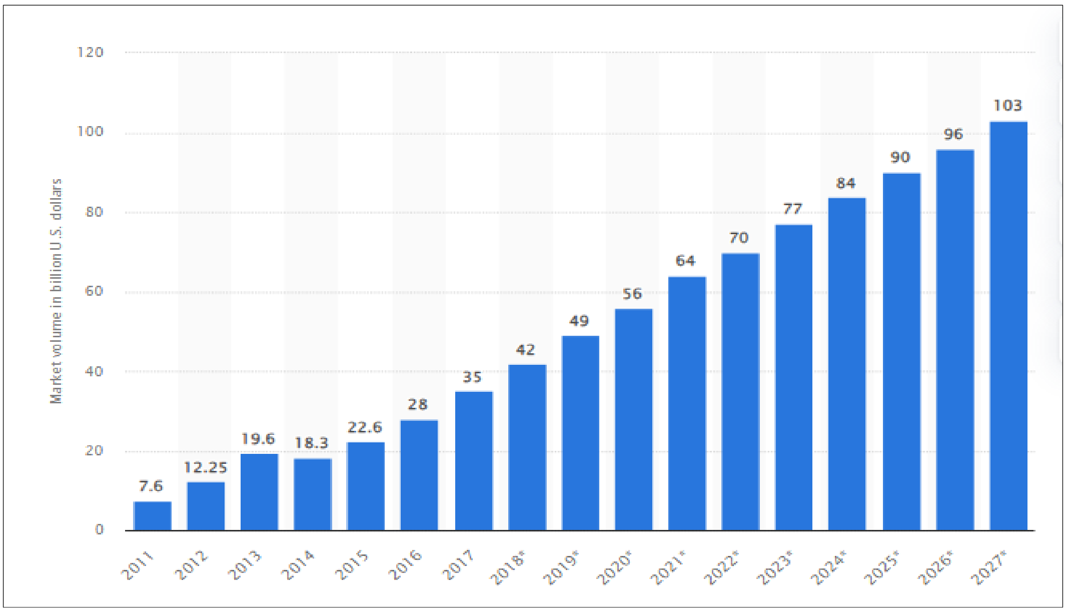

Figure No 1: Big data market size revenue worldwide

The application of streaming data is also provided by many platforms as it saves time and increases fluency in terms of accessing data. As per [3] the Video Streaming market is projected to reach $73.55 in 2023. The increased use of video streaming apps and real-time time trading, gaming, and government policy utilisation is raised. Furthermore, for marketing purposes, the streaming data is also used. The big data clouds are giving the streaming data necessary storage space and increasing the speed of their services.

RESEARCH GAP

The research gap in the performance analysis of streaming data with big data frameworks lies in the limited exploration of scalability, resource optimization, latency analysis, fault tolerance, and benchmarking. Closing this gap is crucial for enabling organizations to efficiently process and analyse large-scale streaming data while maintaining low-latency and reliable operations.

RESEARCH OBJECTIVE

- To examine performance and challenges in streaming data processing with big data frameworks for comprehensive insights.
- To identify factors impacting scalability, resource use, latency, and fault tolerance.
- To offer best practices for high-performance streaming data solution design and optimization.

PROBLEM STATEMENT

A new era of real-time analytics has begun as a result of the increasing use of streaming data processing through big data frameworks, providing unmatched opportunities for prompt insights and decision-making. However, this increase in real-time data processing additionally brings about a number of difficulties that demand careful consideration [7]. The main issue is how to maximise streaming data application performance while coping with rising data volumes, preserving low latency, and ensuring fault tolerance. According to [9], there are privacy violations, risks of error, risks of abuse among data holders, exacerbation of not equal economics in the market, weakening of economy based on the competition and restrictions of civil rights through interference of human activity all are the threat of using big data.

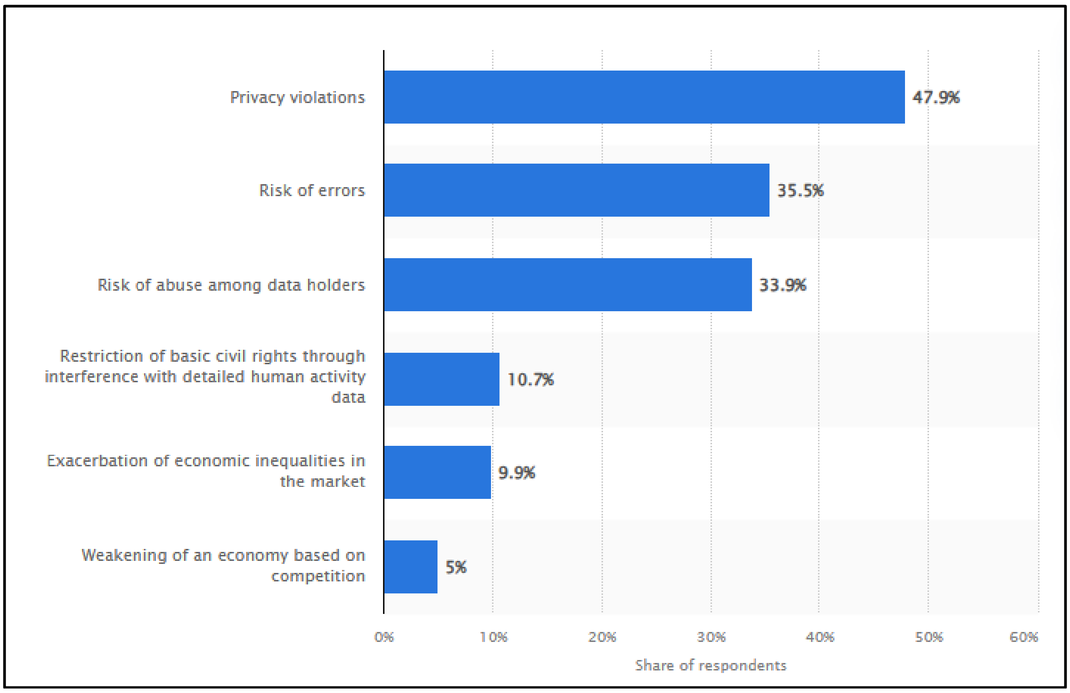

Figure 2: Threats of using Big Data in companies in Poland in 2022

CONCEPTUAL MODEL

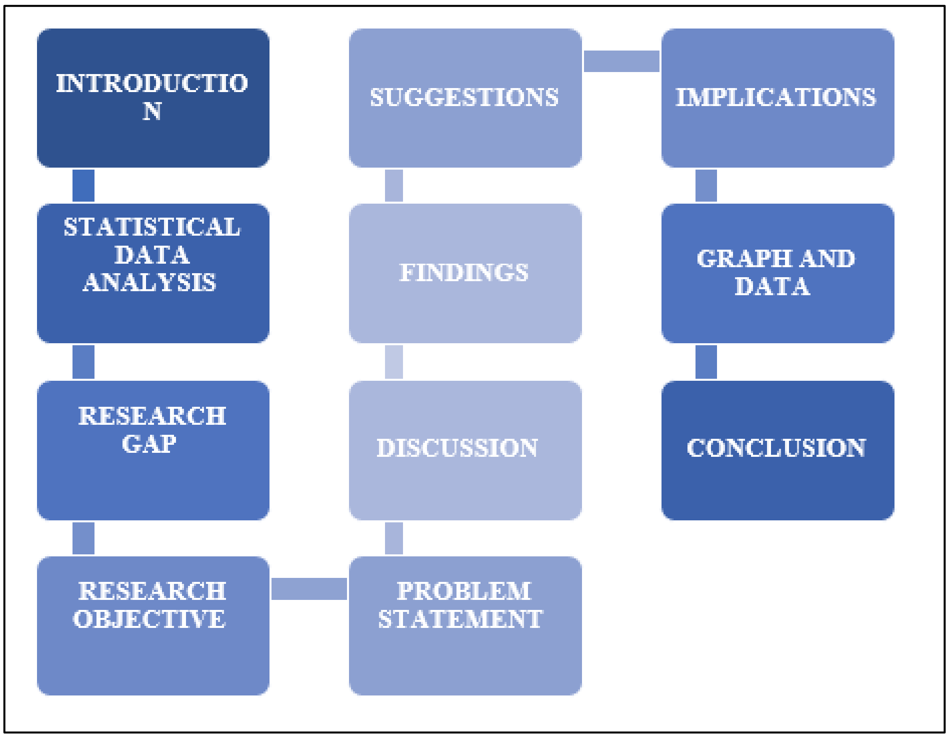

Figure No. 3: article framework

DISCUSSION

This study emphasises the critical importance of monitoring and stream data pipeline optimisation. The need to ensure effective and trustworthy real-time processing becomes crucial as data volumes from sources like Internet of Things devices, online communities, and sensors continue to increase. This study emphasises the importance of a comprehensive strategy for performance analysis. It emphasises how crucial it is to take multiple factors into account when designing efficient streaming data solutions, such as data volume, event time processing, and latency of the network. Effective resource management and allocation are crucial factors in determining the performance of streaming data [6]. The need is for dynamic resource optimisation, especially in cloud-based environments with high resource costs. Reduced operational costs are achieved through efficient management of CPU, memory, and storage resources, which also guarantees consistent high performance.

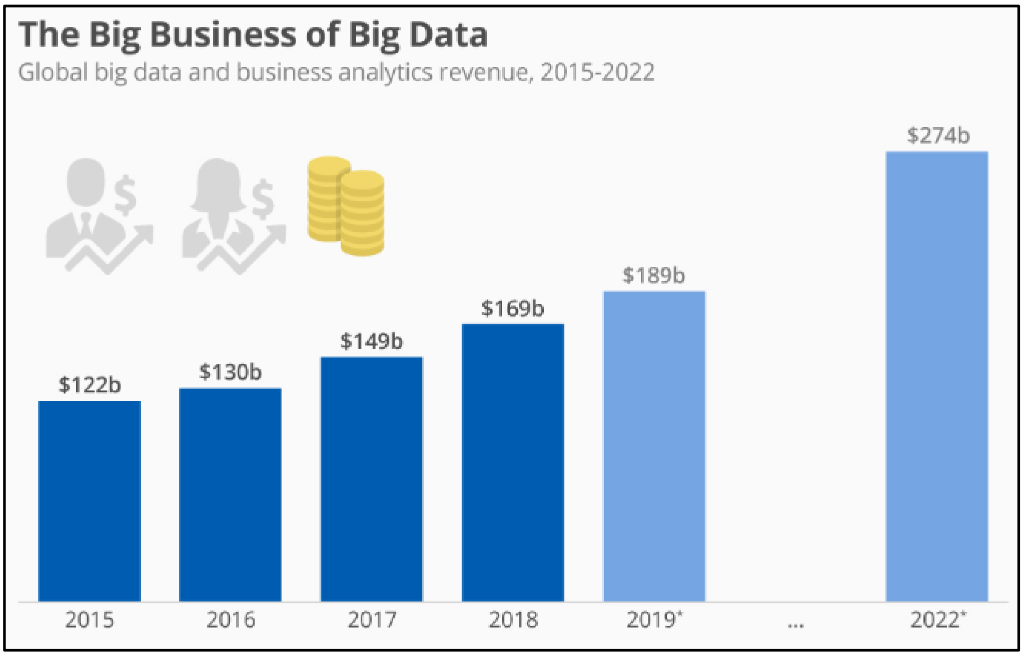

Figure 4: The Big Business of Big Data

In many different industries, performance analysis of streaming data with big data frameworks is crucial. Applications that depend on effective streaming data processing include real-time analytics, fraud detection, monitoring, personalization in e-commerce, and content recommendation. According to [8], the market for big data was $189 billion in 2019 but increased to $247 billion in 2022, showing the rising demand for big data adaptation. Poor performance can result in missed opportunities, delays in decision-making, and higher operational costs. Real-time optimisation is a constant requirement for performance analysis in streaming data with big data frameworks [7]. Streaming data needs to be analysed as it comes in, frequently in under a millisecond, unlike batch processing, where data is processed in set chunks. Since data processing must be done in real time, continuous monitoring, adaptive resource management, and dynamic scaling are required. This presents both unique challenges and opportunities for the field of data engineering and analytics. Nevertheless, in the fields of real-time analysis, predictive analytics, improving personalisation, connecting the IoT and smart services, supply chain optimisation and operational efficiency streaming data has contributed.

FINDINGS

Scalability is still a big issue when working with streaming data, especially as data volumes increase quickly. Larger workloads frequently require big data frameworks to adapt without compromising performance. The significance of dynamic resource allocation in maximising the effectiveness and affordability of processing streaming data. The effective use of CPU, memory, and storage resources could be emphasised as a crucial element. As per the view of [4] Performance can be improved by allocating more computing resources, such as CPUs and memory, but this has a higher operational cost. This is because more resources can handle larger workloads and reduce latency. The growing trend is related to real-time data processing and has a developing trend, which is the use of edge computing for real-time data processing. Streaming data processing solutions must be continuously improved over time, which requires regular performance evaluations and cost analyses.

SUGGESTIONS

The cost-performance trade-offs involved in processing streaming data. Organisations might have to find a balance between operational costs and resource provisioning. To handle the constantly increasing volume of streaming data without sacrificing performance, organisations must prioritise the development of horizontal scaling strategies, load balancing systems, and auto-scaling capabilities. Organisations should choose big data frameworks based on the available data, this ensures the framework of choice satisfies their unique streaming data needs and offers the desired level of performance and scalability.

IMPLICATIONS

The use of big data frameworks in streaming data enhances real-time decision-making, Organisations can optimise their streaming data lines for real-time processing by diving deep into performance analysis. This results in quicker decision-making abilities that let businesses react quickly to shifting circumstances, spot anomalies, and take advantage of opportunities more successfully. Creating and implementing resource optimisation strategies that match resource provisioning with actual workload requirements [6]. This aids organisations in finding a balance between cost-effectiveness and high performance. The development of latency reduction strategies is facilitated by an understanding of the complexities of latency in real-time data processing. This has implications for sectors like finance, healthcare, and e-commerce where low-latency processing is essential. It enables businesses to provide services that are quicker and more responsive.

The implications of this article go right to the basis of contemporary data-driven organisations. They emphasise the significance of stream data performance optimisation, taking scalability, resource efficiency, low latency, and reliability into account. By putting these implications into practice, organisations can fully utilise real-time data analytics, maintain their competitiveness, and make quicker and more informed decisions. Building real-time data pipelines and streaming applications primarily uses the open-source messaging platform Kafka [6]. Powerful distributed computation engines, distributed databases, cloud storage, and virtualization technologies are the benefits of cloud computing technology.

GRAPH AND DATA

In the past few years, the popularity of streaming data increased as a result of the adaptation of technologies that use streaming data more. As observed by [5] audio streaming has increased from 11 per cent to 13 per cent. On the other hand, video streaming is raised from 44 per cent to 47 per cent. The rise in the popularity of streaming data signifies that streaming data is gaining popularity among the systems. Products like music and video streaming apps are gaining profit with the well-developed application.

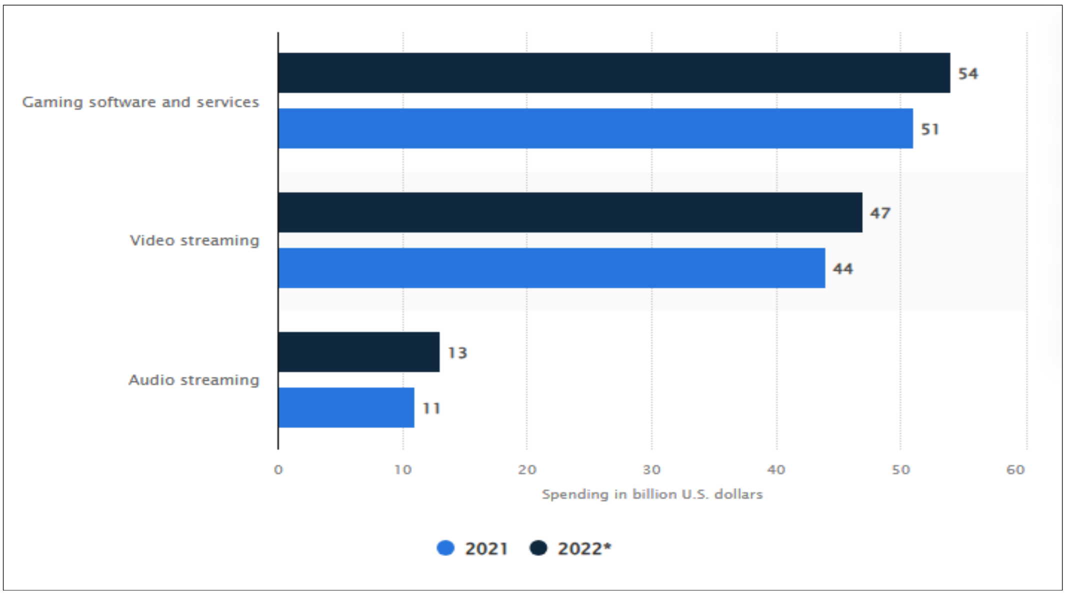

Figure No. 5: Consumer spending on video games, video and audio streaming

CONCLUSION

In conclusion, the article emphasises how crucial effective real-time data processing is in today's data-driven environment. It highlights the complex interplay between various elements, including scalability, resource optimisation, latency reduction, fault tolerance, and technology choice, in achieving high-performance streaming data solutions. The implications of this research extend to businesses looking to capitalise on the power of streaming data, highlighting the necessity of cost-effective resource management, informed decision-making, and continuous improvement. The knowledge gained from this study serves as a road map for organisations to travel through the challenging landscape of streaming data analytics, allowing them to unlock timely insights, improve competitiveness, and provide more responsive services as the volume and velocity of data continue to increase.

References

[1]. Petroc. T., (2022), Big data market size revenue forecast worldwide from 2011 to 2027. Available at: https://www.statista.com/statistics/254266/global-big-data-market-forecast/ [accessed on: 13 September 2023]

[2]. Petroc. T., (2023), Big data - statistics & facts. Available at: https://www.statista.com/topics/1464/big-data/ [accessed on: 13 September 2023]

[3]. Statista Market Insights, (2023), Video Streaming (SVoD) - Worldwide. Available at: https://www.statista.com/outlook/dmo/digital-media/video-on-demand/video-streaming-svod/worldwide#revenue [accessed on: 13 September 2023]

[4]. Fé, I., Matos, R., Dantas, J., Melo, C., Nguyen, T.A., Min, D., Choi, E., Silva, F.A. and Maciel, P.R.M., (2022). Performance-Cost Trade-Off in Auto-Scaling Mechanisms for Cloud Computing. Sensors, 22(3), p.1221.

[5]. Stoll.J., (2023), Consumer spending on video games, video and audio streaming in the United States in 2021 and 2022. Available at: https://www.statista.com/statistics/1303900/gaming-and-audio-video-streaming-spendings-usa/ [accessed on: 13 September 2023]

[6]. Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big data service architecture: a survey. Journal of Internet Technology, 21(2), 393-405.

[7]. Gomes, H. M., Read, J., Bifet, A., Barddal, J. P., & Gama, J. (2019). Machine learning for streaming data: state of the art, challenges, and opportunities. ACM SIGKDD Explorations Newsletter, 21(2), 6-22.

[8] Feldman.S., (2019), The Big Business of Big Data. Available at: https://www.statista.com/chart/18328/big-data-business-analytics-revenue/ [accessed on: 13 September 2023]

[9] Sas. A., (2022), Threats of using Big Data in companies in Poland in 2022. Available at: https://www.statista.com/statistics/1370593/poland-big-data-threats-in-companies/ [accessed on: 13 September 2023]