A Scalable Network Analytics Structure & Procedure for Real-time Processing of Enormous Data

Dr.P.Chellammal,
Professor, Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Trichy, Tamilnadu
R.Shariff Nisha,
Assistant Professor, Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Trichy, Tamilnadu
T.Vency Stephisia,
Assistant Professor, Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Trichy, Tamilnadu
S.Harthy Ruby Priya,
Assistant Professor, Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Trichy, Tamilnadu

Abstract

This research presents a network analytics structure and procedure designed to support real-time processing of large-scale data. The structure consists of multiple data adaptive nodes and data analytic clusters arranged in a distributed network. The data analytic nodes within the clusters employ a peer-to-peer networking mode and a load balancing mechanism, enabling the clusters to dynamically expand and contract as needed. The structure utilizes a streamlined, flow-based analytic processing approach among the data analytic nodes, facilitated by an incident mechanism. By leveraging this network analytics structure, real-time analysis and processing of enormous network data, including fault monitoring, statistics, troubleshooting, and diagnosis, can be achieved. The structure allows for fine-grained analysis of network data and supports dynamic expansion of structure functionalities, allowing users to define their own analytic requirements. Additionally, the structure's distributed architecture reduces reliance on individual hardware performance, enabling efficient processing of complex logic for network data analysis. The structure and procedure support various types of processing logics, thereby reducing the expertise required from developers.

Keywords: Network analytics, Real-time data processing, Distributed structure, Scalability, Peer-to-peer networking, Load balancing, Fine-grained analysis
DOI Number: 10.48047/nq.2020.18.8.nq20226 **NeuroQuantology 2020;18(8):197-202**

Introduction

In today's digital age, the proliferation of network-connected devices and the exponential growth of data have presented new challenges and opportunities for organizations. The ability to harness and analyze enormous amounts of network data in real-time has become crucial for enhancing network performance, detecting anomalies, and optimizing operational efficiency. As a result, there is a growing need for advanced network analytics structures and procedures that can handle the complexities of processing and analyzing vast volumes of data in a timely and efficient manner.[1]

This research focuses on developing a network analytic structure and procedure that supports the real-time processing of enormous data. The objective is to create a scalable and distributed architecture capable

of handling diverse network operations, such as fault monitoring, statistics, troubleshooting, and diagnosis. By leveraging cutting-edge technologies and innovative approaches, this research aims to provide organizations with the tools and capabilities to extract valuable insights from their network data in real-time.[2]The proposed network analytics structure is designed to address the limitations of traditional analytics approaches, which often struggle to handle the sheer volume and velocity of data generated in today's networks.

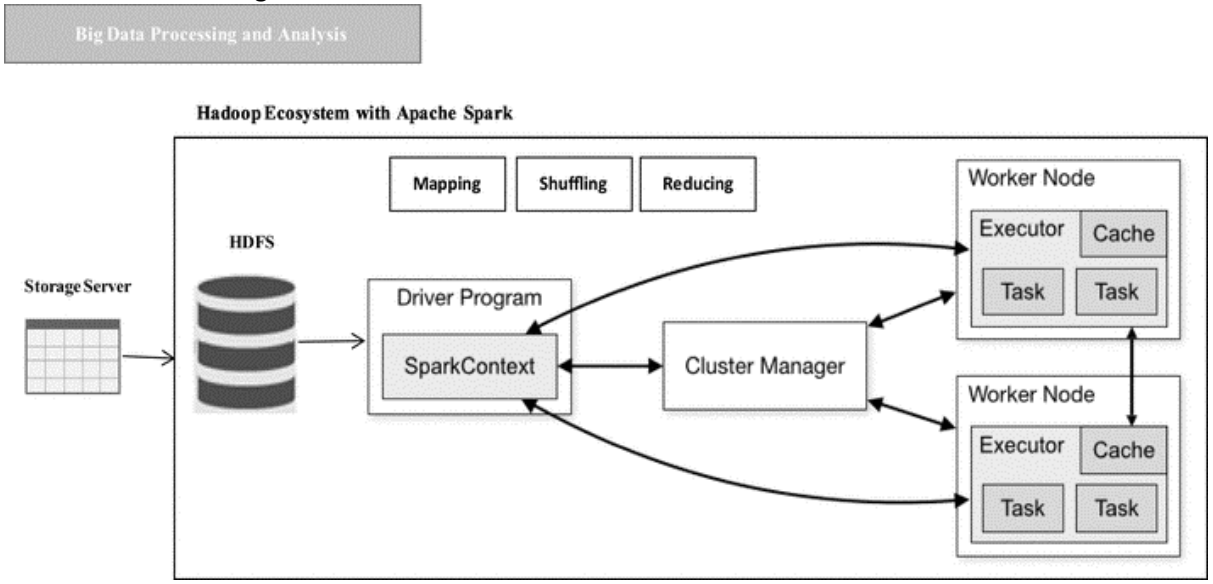

Figure 1. Big Data Processing and Analysis Framework

The structure comprises a network of interconnected data adaptive nodes and data analytic clusters, which are arranged in a distributed fashion. This distributed architecture allows for efficient data processing and analysis across multiple nodes, enabling real-time decision-making and action.One of the key features of the structure is the utilization of a peer-to-peer networking mode and a load balancing mechanism within the data analytic nodes.[3] This enables the structure to dynamically expand and contract its resources based on the workload and demand. By distributing the computational tasks across multiple nodes, the structure can achieve high scalability and adaptability to varying data processing requirements.

The structure also employs a flow-line type analytic processing process facilitated by an incident mechanism. This streamlined approach ensures efficient data flow and enables seamless collaboration among the data analytic nodes. It allows for the real-time analysis of network data, capturing insights and patterns as they emerge, and facilitating proactive decision-making and problem resolution.Moreover, the structure supports fine-grained analysis of network data, enabling organizations to delve deeper into the intricacies of their networks.[4] By considering multiple characteristics and parameters, such as network performance metrics, traffic patterns, and device behavior, the structure can provide comprehensive insights into the network's functioning and identify potential issues or opportunities.

In addition to real-time processing and fine-grained analysis, the structure also offers a self-defined analytic capability. This means that users can customize and define their own analytic requirements based on the specific types of data they want to detect and analyze. This flexibility allows organizations to tailor the structure to their unique needs and extract meaningful insights that are relevant to their operations.By leveraging the distributed nature of the structure, the proposed network analytics approach reduces reliance on single hardware performance and mitigates the risk of bottlenecks.[5] The distributed architecture allows for efficient processing of complex logic and ensures

structure robustness and scalability.The research presented in this paper aims to advance the field of network analytics by developing an advanced structure and procedure for real-time processing of enormous data. By combining scalability, fine-grained analysis, and self-defined analytics, the proposed structure empowers organizations to extract valuable insights from their network data, optimize network performance, and enhance operational efficiency.[6] The subsequent sections of this research will delve deeper into the structure architecture, the procedureology employed, and the experimental results, highlighting the advantages and benefits of the proposed network analytics approach.

Related Work

In recent years, the Internet has experienced rapid development and expansion in World. With the continuous growth of network size and the emergence of new technologies, the internet has become a catalyst for economic development, social progress, and improvements in people's living standards. The impact of networking has increased significantly, facilitating quick and efficient communication and information exchange.Over the past decade, the development of the internet in World has exhibited remarkable vitality. Significant progress has been made in terms of internet infrastructure, operational modes, and the variety and quality of services offered. The application landscape of the internet in World has become increasingly diverse, including platforms such as blogs, social networking services, and Web2.0-based social interaction platforms like Social Networking Service (SNS).[1]

New businesses, such as classified information services, point-to-point (P2P) interconnection techniques, and innovative management models, have emerged rapidly. This trend indicates a shift towards a more personalized and interactive internet environment.

Simultaneously, the internet has started to penetrate traditional industries and service sectors. Personal websites have flourished, and e-business applications have expanded from enterprise-focused to individual-focused, extending to traditional forms of businesses as well. This infiltration of the internet into traditional industries provides ample opportunities for development, playing a significant role in fostering an innovation-oriented country. The management of intranets within enterprises has garnered increasing attention due to its impact on their operations.In recent years, the intranet architecture in World has gradually adopted a framework consisting of routers, firewalls, switches, and servers. While the positive effects of the internet are becoming increasingly apparent, new problems and vulnerabilities continue to emerge, presenting challenges for network management and security.[7]The problems and threats encountered in networks can be categorized into four main types: network congestion, unknown causes leading to packet loss, network equipment failures resulting in unavailable network services, unauthorized access, and routing errors.

To effectively address different types of security threats, it is necessary to employ various monitoring and management measures.Conventional network surveillance software, such as NETSCOUT's Sniffer Adaptive Application Analyzer and the network packet analysis software WireShark, provide a range of monitoring and alarm feedback functionalities. However, these applications have certain limitations. As network bandwidth increases and network traffic data grows, legacy network monitoring software reaches its functional limits, making it incapable of real-time monitoring tasks.

Additionally, legacy network monitoring software suffers from three key weaknesses. First, it supports a limited range of network data analysis types, typically focusing on universal classes of network analyzing data. This limitation hampers the ability to meet the diverse and specific network data analysis demands of enterprise networks. Furthermore, these software solutions often exhibit inaccuracies in analysis and positioning.[3] For example, the Sniffer software can only display the flow of network data between the source and destination data

terminals, rather than providing a comprehensive view of all data paths. Real-time data analysis is another shortcoming of legacy network monitoring software, as it primarily focuses on data capture but lacks the ability to process data in real-time.

Three prior art approaches to network data analysis will now be introduced, along with their associated shortcomings:

Scheme One: Network data analysis using the Simple Network Management Protocol (SNMP) statistical data provided by switches. This approach involves using SNMP to obtain statistical information from switches and conducting macroscopic analysis of network conditions and operations. While this scheme enables the analysis of network congestion and macroscopic situations like packet loss, it heavily relies on switch-provided information and functions, offering limited flexibility for users. Additionally, this scheme does not facilitate the analysis of specific data streams or provide precise monitoring and control of particular data.

Scheme Two: Network data analysis using network packet capturing software, commonly known as Sniffer. This procedure involves capturing packets using Sniffer software and allowing users to search for and analyze the required data by setting filterrules. Although this approach allows for the monitoring of specific data streams, its analytical model is fixed and does not support user-defined analysis types. Moreover, Sniffer software is limited to desktop environments, resulting in low efficiency and a lack of dynamic scalability. It cannot adequately support gigabit networking and requires regular updates to enhance its functionality and performance.[6]

Scheme Three: Network data analysis using Deep Packet Inspection (DPI) technology. This procedure involves deep reading of IP packets, recombining application layer messages, and obtaining comprehensive application content. Based on predefined management strategies, the structure performs shaping operations on data flows. DPI technology primarily focuses on purposes such as traffic management and network security, but it lacks robust network data analysis capabilities. Its implementation often relies on hardware performance, making it less suitable for complex process logic and requiring high technical expertise from developers.[7]In summary, the prior art approaches discussed above face several technical challenges that need to be addressed. The key issues include designing a mass data real-time processing platform that meets the requirements of a distributed structure with dynamic expansion capabilities. Additionally, the flexible and efficient management of the structure's internal components in a distributed manner needs to be tackled. Furthermore, allowing users to define their own data analysis types and supporting a wide range of network data analysis requirements are crucial objectives. These challenges are of great interest to professionals in the industry, and finding effective solutions will significantly advance the field of network data analysis.

Research Objective

The objective of this research is to develop a network analytics structure and procedure capable of real-time processing of enormous data. The focus is on creating a scalable and distributed architecture that supports dynamic expansion of structure functions and enables fine-grained analysis of network data. The research aims to provide real-time analysis and processing capabilities for various network operations, such as fault monitoring, statistics, troubleshooting, and diagnosis. Additionally, the goal is to empower users to define their own analytic requirements and reduce the expertise required from developers.

A Scalable Network Analytics Structure & Procedure for Real-time Processing of Enormous Data

The network analysis structure supports real-time processing of large amounts of data. It consists of multiple data adaptation nodes and data analysis clusters. The data adaptation nodes gather data from the external network and send it to the data analysis clusters for real-time analysis. They can convert the data into a format that the analysis nodes in the data analysis cluster can

understand and communicate with them effectively. The workload is balanced by distributing the events to the analysis nodes in the cluster. The results of the data analysis process are then sent back to the data adaptation nodes and can be exported to external structures, such as databases, logs, or graphical interfaces, for storage, recording, or display purposes. The input to the data adaptation nodes includes all types of data from the external network, analysis requests, intermediate events, or final output events. The output is adapted to suit the specific needs of databases, logs, graphical interfaces, or network communication in various formats.The data analysis cluster is the core component of the structure, responsible for performing real-time monitoring, analysis, aggregation, investigation, and diagnosis of network data. It consists of multiple decentralized data analysis nodes connected in a network. These nodes handle tasks such as monitoring network malfunctions, analyzing network operations, performing calculations, and diagnosing issues. The results of the analysis can be sent back to the data adaptation nodes or stored in databases, logs, and graphical interfaces for further processing. The data analysis nodes support a peer-to-peer networking mode and a load-balancing mechanism, allowing the cluster to adjust its processing capacity by adding or removing analysis nodes as needed. The analysis and processing of data among the nodes are carried out through event-based communication in a pipeline structure.

Conclusion

In conclusion, this research has presented a network analytics structure and procedure that supports real-time processing of enormous data. The structure's distributed architecture, incorporating data adaptive nodes and data analytic clusters, allows for dynamic expansion and contraction to accommodate varying data processing needs. The utilization of a flow-based analytic processing approach among the data analytic nodes enables efficient and streamlined processing of complex network data logics. The structure's support for fine-grained analysis and its ability to handle various types of processing logics make it a versatile solution for real-time network data analysis. By reducing the dependency on individual hardware performance, the structure offers scalability and robustness. Overall, this research contributes to the advancement of network analytics by providing a flexible and efficient structure for processing enormous data in real time.

Reference

1. H. Hu, Y. Wen, T. -S. Chua and X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial," in IEEE Access, vol. 2, pp. 652-687, 2014, doi: 10.1109/ACCESS.2014.2332453.
2. Singh, K., Guntuku, S. C., Thakur, A., &Hota, C. (2014). Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests. Information Sciences, 278, 488-497. https://doi.org/10.1016/j.ins.2014.03.066
3. AriyaluranHabeeb, R. A., Nasaruddin, F., Gani, A., Targio Hashem, I. A., Ahmed, E., & Imran, M. (2019). Real-time big data processing for anomaly detection: A Survey. International Journal of Information Management, 45, 289-307. https://doi.org/10.1016/j.ijinfomgt.2018.08.006
4. L. Cui, F. R. Yu and Q. Yan, "When big data meets software-defined networking: SDN for big data and big data for SDN," in IEEE Network, vol. 30, no. 1, pp. 58-65, January-February 2016, doi: 10.1109/MNET.2016.7389832.
5. Liu, Z., Jiang, B., &Heer, J. (2013). ImMens: Real-time Visual Querying of Big Data. Computer Graphics Forum, 32(3pt4), 421-430. https://doi.org/10.1111/cgf.12129
6. Rathore, M. M., Paul, A., Hong, W., Seo, H., Awan, I., & Saeed, S. (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. Sustainable Cities and Society, 40, 600-610.

https://doi.org/10.1016/j.scs.2017.12.022

7. H. Liu, Y. -S. Ong, X. Shen and J. Cai, "When Gaussian Process Meets Big Data: A Review of Scalable GPs," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4405-4423, Nov. 2020, doi: 10.1109/TNNLS.2019.2957109.
8. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., &Ullah Khan, S. (2014). The rise of “big data” on cloud computing: Review and open research issues. Information Systems, 47, 98-115. https://doi.org/10.1016/j.is.2014.07.006
9. Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., &Vasilakos, A. V. (2016). Big data: From beginning to future. International Journal of Information Management, 36(6), 1231-1247. https://doi.org/10.1016/j.ijinfomgt.2016.07.009