ISSN: 2688-8645

Research on the Architecture of Maritime Regulatory Data Platform

Enze Wu, Jin Tao, Taizhi Lv

College of Information Engineering, Jiangsu Maritime Institute, Nanjing 211170, China

Abstract

The maritime regulatory data platform serves as the brain of the smart maritime regulatory platform. It collects data on navigation resources and port & waterway resources, unifies data norms and standards, and achieves comprehensive data resource management covering all elements including ships, crew members, companies, ports, waterways, and pilotage. Through data aggregation, mining, and analysis, it enables intelligent judgment, precise profiling, scientific decision-making, and multi-dimensional panoramic display. This platform provides maritime regulatory departments with scientific and reasonable decision-making suggestions, helping them better formulate and implement regulatory policies.

Keywords

Maritime Regulation; Data Analysis; Lambda Architecture; Data Warehouse; Flume.

1. Introduction

Driven by global economic integration and the wave of digitalization, the shipping industry is undergoing unprecedented changes [1]. With the continuous growth of international trade, the shipping industry, as a key link in the international logistics chain, is becoming increasingly important. However, traditional maritime regulation and shipping industry operation models face numerous challenges, such as information silos, inefficiencies, and safety hazards. To address these challenges, data empowerment has emerged as a crucial force driving the transformation and upgrading of maritime regulation and the shipping industry [2-3]. The rapid development of advanced technologies such as big data, artificial intelligence, and the Internet of Things provides a solid foundation for data empowerment. By integrating and analyzing various data resources, maritime regulatory authorities and shipping companies can more precisely grasp information on ship operation status, waterway traffic conditions, and market demand changes, enabling them to make more scientific and reasonable decisions.

The significance of data empowerment in maritime regulation and the shipping industry is profound. It helps enhance the efficiency and accuracy of maritime regulation. Through real-time monitoring and data analysis, maritime authorities can promptly identify and address safety hazards, reducing the incidence of maritime traffic accidents. Data empowerment also assists maritime authorities in more accurately understanding ship operation status and market demand changes, providing a basis for formulating scientific and reasonable regulatory policies. Furthermore, data empowerment contributes to promoting the high-quality development of the shipping industry. By integrating and analyzing various data resources, shipping companies can more precisely formulate shipping routes and transportation plans, optimize resource allocation, and improve transportation efficiency and service quality. Additionally, data empowerment facilitates the deep integration of the shipping industry with other industries, driving the industry towards intelligence, greenness, and service-oriented development.

ISSN: 2688-8645

The maritime regulatory data platform, as the core brain of the smart maritime regulatory platform, holds immense importance. By extensively collecting data on navigation resources and port & waterway resources, the platform constructs a vast and comprehensive data warehouse. The all-encompassing data resource management approach provides maritime regulation with unprecedented depth and breadth of information, serving as an important cornerstone for achieving intelligent regulation. Based on data aggregation, the maritime regulatory data platform leverages advanced data mining and analysis technologies to deeply process these massive amounts of data. Through intelligent algorithms, the platform can automatically identify complex correlations between data, revealing hidden patterns and trends. Multi-dimensional panoramic displays present complex data analysis results in an easily understandable format through intuitive charts, dynamic maps, and other forms, allowing regulatory authorities to quickly grasp the overall situation in the maritime field, identify and resolve issues promptly. Whether for daily regulation, emergency response, or long-term planning, the maritime regulatory data platform provides powerful support, assisting maritime regulatory authorities in improving work efficiency, ensuring water transportation safety, and promoting the sustainable development of the shipping industry.

2. System Architecture

As illustrated in Figure 1, the production environment of the smart maritime regulatory platform is a highly integrated and powerful system architecture, primarily consisting of four core components: the application layer, data acquisition layer, data resource layer, and data computing layer. Each layer plays an indispensable role, collectively ensuring the stable operation and efficient service of the platform. The data computing layer, serving as the central brain of the entire smart maritime regulatory platform, is the core focus of the maritime regulatory data platform discussed in depth in this paper. This layer is responsible for efficient processing and in-depth analysis of shipping big data. By utilizing advanced big data processing technologies and algorithmic models, it cleans, integrates, and mines massive amounts of data, extracting valuable information and insights to provide strong data support for maritime regulation and shipping decisions. It not only enhances the speed and quality of data processing but also maximizes the utilization of data value. The application layer is the direct interface of the platform for users, hosting the informational operation tasks of multiple business systems such as the Intelligent Ship Control System, Comprehensive Ship Law Enforcement Management System, and Ship Shore Power Management System. These systems, through user-friendly interfaces and efficient operational processes, enable intelligent management of the entire process of ship navigation, berthing, and operations, as well as rapid response and handling of ship violations and irregularities. Simultaneously, the application layer is responsible for visualizing shipping big data in various forms such as charts, reports, and maps, making complex data information intuitive and easy to understand, facilitating users to quickly grasp maritime dynamics and make accurate decisions. The data acquisition layer is the foundation for the platform to obtain data. It is responsible for real-time acquisition of multi-source heterogeneous data such as ship dynamics, waterway conditions, and meteorological information from AIS (Automatic Identification System), VTS (Vessel Traffic Service) systems, and various business systems, providing comprehensive, accurate, and timely data input for the platform. The data resource layer bears the responsibility of storing business data. By constructing database clusters, data cache clusters, and file storage clusters, it achieves efficient data organization, rapid access, and secure storage, providing solid data support for the data computing layer and application layer.

ISSN: 2688-8645

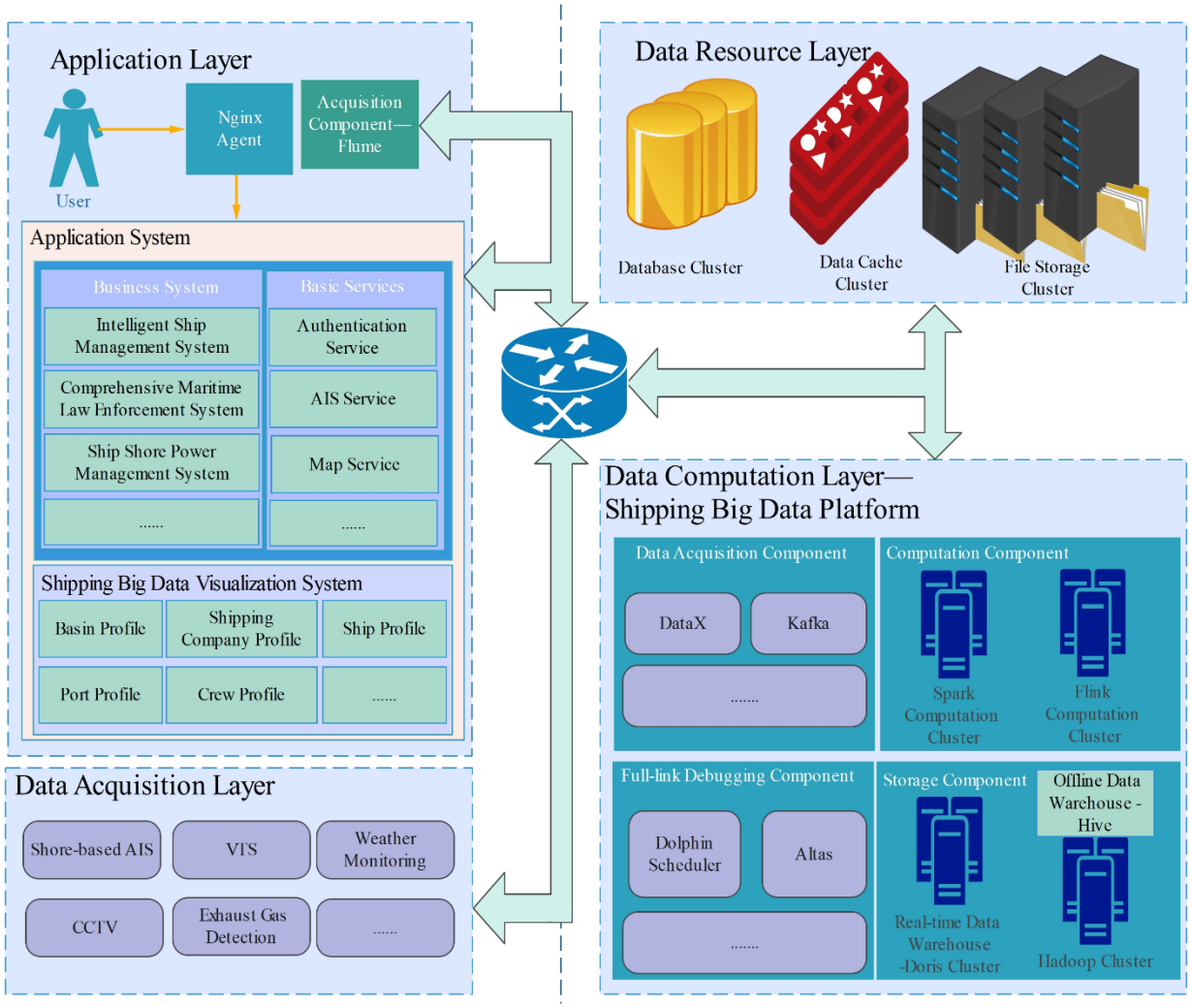

Fig 1. The platform architecture

The entire platform is deployed with over 30 physical hosts and has virtualized more than 200 private cloud hosts, divided into internal and external network sections. The internal network includes the Data Resource Layer and the Maritime Regulatory Data Platform, while the external network comprises the Application Layer and the Data Acquisition Layer. The hardware scale will increase with the expansion of business and data.

3. Data Processing

The maritime regulatory data platform is based on the Lambda architecture and consists of data acquisition components, computing components, storage components, and full-link debugging components. The data acquisition components import data from the Data Acquisition Layer and the Data Resource Layer. The data acquisition involved includes importing data from databases, importing data from log files, and collecting data based on message queues. The Data Computing Layer completes both offline and real-time data computations, with the results stored in databases. Offline data is processed based on the Spark computing cluster, while real-time data is handled by the Flink computing cluster. The data storage components enable the storage of massive amounts of data, divided into real-time data warehouses and offline data warehouses. The full-link debugging components manage the task scheduling of data throughout the entire process, from acquisition and storage to computation.

ISSN: 2688-8645

3.1. Data Acquisition Components

In the data acquisition layer of the smart maritime regulatory platform, the data acquisition components are divided into three key modules to ensure comprehensive and efficient data handling. Firstly, for data acquisition from business databases, the platform selects DataX, a powerful data migration tool. DataX efficiently transfers data from business databases to the big data platform, achieving seamless data integration and laying a solid foundation for subsequent data analysis [4]. Secondly, for log file acquisition, the platform adopts Flume, a professional log collection tool. Flume can capture and process various log files in real-time, converting them into structured data for subsequent analysis and mining [5]. Finally, to obtain richer shipping-related data, the platform also uses collection programs such as the Internet of Ships and web crawlers to acquire key data such as AIS (Automatic Identification System) data, ship meteorological information, and ship status in real-time. This data is aggregated and distributed through the Kafka message queue, ultimately flowing into the big data platform, providing comprehensive and timely data support for maritime regulation and shipping decisions.

3.2. Data Computing Components

In the data computing layer of the smart maritime regulatory platform, the data computing components are divided into real-time computing and offline computing modules to meet data processing needs in different scenarios. The offline computing module is primarily built on the Spark computing cluster. It periodically executes preset data analysis programs, deeply mining the massive shipping data stored in the data warehouse. Through complex data processing and analysis algorithms, the offline computing module can uncover deep-level patterns and trends behind ship navigation behavior, providing maritime regulatory departments with valuable analysis results and insights such as ship traffic analysis, route optimization suggestions, and safety risk warnings. These analysis results and insights are significant for maritime regulatory departments to formulate scientific and reasonable policies and improve regulatory efficiency.

The real-time computing module, on the other hand, relies on the Flink computing cluster. It focuses on receiving and processing various data generated during ship navigation in real-time, including critical data such as navigation status, location information, and cargo status. Through efficient stream data processing techniques, the real-time computing module can parse, compute, and analyze this data in real-time, providing maritime regulatory departments and shipping companies with timely and accurate decision support. This not only helps ensure the safe navigation of ships but also plays a crucial role in cargo scheduling, energy efficiency management, and other aspects, promoting the intelligent and efficient development of the shipping industry.

The organic combination of offline and real-time computing enables the smart maritime regulatory platform to simultaneously meet the needs for deep mining of historical data and rapid response to real-time data, providing strong data support for the high-quality development of maritime regulation and the shipping industry.

3.3. Data Storage Components

The data storage components are divided into real-time data warehouses and offline data warehouses to ensure efficient and orderly management and utilization of various types of data. The real-time data warehouse is built on the advanced Doris cluster, which focuses on storing real-time ship navigation data. The Doris cluster can seamlessly receive real-time data streams processed by the Flink computing cluster and efficiently store them in the real-time data warehouse. Thanks to Doris's vectorized query engine technology, data query performance is significantly improved [6]. Maritime regulatory departments can quickly perform real-time queries and analyses on ship navigation data, obtaining the latest ship dynamics and navigation

ISSN: 2688-8645

statuses in a timely manner to provide strong support for safety regulation and emergency response.

The offline data warehouse, on the other hand, relies on Hive and is primarily responsible for storing massive amounts of offline ship navigation data. Hive's distributed storage and computing capabilities enable the offline data warehouse to easily handle the storage and processing needs of large-scale data [7]. By applying data mining techniques, maritime regulatory departments can deeply analyze these offline data, revealing patterns and trends in navigation modes, ship performance, and other key indicators. These analysis results not only help identify potential areas for management optimization but also provide new ideas and directions for the intelligent development of the shipping industry.

The collaboration between the real-time data warehouse and the offline data warehouse enables the smart maritime regulatory platform to comprehensively and efficiently manage various ship navigation data, providing maritime regulatory departments and shipping companies with more precise and timely data support, and contributing to the high-quality development of the shipping industry.

3.4. Full-Link Debugging Components

In the smart maritime regulatory platform, the full-link debugging components are crucial for ensuring smooth data flow. These components, based on advanced tools such as Dolphin Schedule, comprehensively monitor the entire link of shipping data flow, from acquisition, storage, processing to analysis. They can track the data flow status in real-time, ensuring efficient and accurate data transmission between various stages. At the same time, the full-link debugging components have powerful problem detection and resolution capabilities, enabling them to quickly locate and address any anomalies or bottlenecks in the data flow process, thereby ensuring the stability and reliability of the entire data link.

4. Conclusion

As the core engine of the smart maritime regulatory platform, the maritime regulatory data platform fully integrates diverse data such as navigation resources and port and shipping resources, and establishes unified data specifications and standards, achieving comprehensive management of all-element data including ships, crew, companies, ports, waterways, and pilotage. Through efficient data aggregation and in-depth mining and analysis, the platform provides powerful functions such as intelligent analysis, precise profiling, scientific decision-making, and multi-dimensional panoramic display, delivering scientific and precise decision-making bases to maritime regulatory departments and assisting them in formulating and implementing more efficient regulatory policies.

The maritime regulatory data platform will continue to undergo optimization and upgrades, further expanding the breadth and depth of data acquisition and enhancing the real-time and accuracy of data processing. At the same time, the platform will deeply integrate cutting-edge technologies such as artificial intelligence and the Internet of Things, driving maritime regulation towards a more intelligent and refined direction.

Acknowledgments

This work was financially supported by the funding of the Research Program of Jiangsu Maritime Institute(2022ZKyb02), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (22KJB580002), and the Excellent Teaching Team for QingLan Project of the Jiangsu Higher Education Institutions of China (Big Data Technology Teaching Team with Shipping Characteristic).

ISSN: 2688-8645

References

[1] Oloruntobi, Olakunle, et al. "Sustainable transition towards greener and cleaner seaborne shipping industry: Challenges and opportunities." Cleaner Engineering and Technology 13 (2023): 100628.

[2] Mallam, Steven C., Salman Nazir, and Amit Sharma. "The human element in future Maritime Operations–perceived impact of autonomous shipping." Ergonomics 63.3 (2020): 334-345.

[3] Kim, Mingyu, et al. "Autonomous shipping and its impact on regulations, technologies, and industries." Journal of International Maritime Safety, Environmental Affairs, and Shipping 4.2 (2020): 17-25.

[4] Coviello, Giuseppe, et al. "DataX: A system for data exchange and transformation of streams." International Symposium on Intelligent and Distributed Computing. Cham: Springer International Publishing, 2021: 319-329.

[5] Wedashwara, Wirarama, et al. "Data storage and modeling system for GPS, Gyro and camera data using apache flume and hadoop map reduce." AIP Conference Proceedings. Vol. 3026. No. 1. AIP Publishing, 2024.

[6] Zou, Yujuan, Peiyi Tang, and Taizhi Lv. "Design and implementation of ship shore power data analysis system based on Doris data warehouse." 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). IEEE, 2022.

[7] Lv, Taizhi, Enze Wu, and Peiyi Tang. "Research on ship manning data analysis based on data warehouse." Third International Conference on Electronic Information Engineering, Big Data, and Computer Technology (EIBDCT 2024). Vol. 13181. SPIE, 2024: 1245-1250.