
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202310.01552

Postprint: Reflections on the Development of Xinhua News Agency Data Exchange Platform

Authors: Zhang Chi

Date: 2023-10-08T00:00:00+00:00

Abstract

The Xinhua News Agency Data Exchange Platform, as a critical technical system of Xinhua News Agency, has been operating stably for over 20 years. In recent years, the development of new media technologies and mobile internet has precipitated substantial transformations in news information production and dissemination paradigms. Consequently, the existing system architecture and business model have become increasingly inadequate to address emerging business requirements. This paper conducts a systematic review and analysis of the platform's current state to identify critical pain points and propose strategic directions for future development and improvement, thereby aligning with technological trends and enhancing service to the news information dissemination enterprise.

Full Text

Preamble

Journal: Media Technology and Business

Title: Evolution and Future Development of Xinhua News Agency's Data Exchange Platform

Keywords: data transmission, data service, unified storage, distributed deployment

Classification: TP39

Document Code: A

ISSN: 1671-0134 (2019) 06-099-02

DOI: 10.19483/j.cnki.11-4653/n.2019.06.029

Author: Xinhua News Agency Technical Bureau

Introduction

Xinhua News Agency's data exchange platform has served as a critical technical system for over two decades. In recent years, the rapid development of new media technologies and mobile internet has fundamentally transformed news production and dissemination methods, rendering the existing system architecture and business model increasingly inadequate for emerging requirements. This paper analyzes the current state of the data exchange platform to identify key pain points and propose future development directions and improvement strategies that align with technological trends, thereby enhancing service capabilities for news and information dissemination.

Currently, Xinhua's headquarters operates dozens of technical systems, including editorial systems, publishing systems, office automation, databases, and Xinhua News Network, in addition to 31 domestic bureaus and 11 overseas bureaus. These components form a massive, hierarchical business network centered at headquarters. As Xinhua's all-media news operations expand, both the number of technical systems and the frequency of information flow within and between these systems—both internal and external—have increased dramatically, creating growing demand for information sharing and exchange across different network domains, systems, and data formats.

Xinhua's communication system, established in the 1990s, has long served as one of the agency's core technical systems, facilitating data exchange between headquarters systems, internal and external networks, domestic and foreign bureaus, and external organizations. The platform processes Xinhua's finished products including text, images, audio, video, and multimedia content, as well as foreign media news and heterogeneous external data. Over the years, continuous business expansion and iterative development have transformed the communication system into a comprehensive and practical data exchange platform that provides foundational data transmission services across the agency and to relevant external organizations.

Around 2010, the world entered the mobile internet era, bringing revolutionary changes to news production and distribution models. As a foundational service provider for Xinhua, the data exchange platform must adapt quickly to these trends by identifying constraints on its own development and transformation.

Current Pain Points

Operational Complexity and Maintenance Pressure: The platform currently employs a one-to-one cold standby approach between primary and backup servers, with dispersed business deployments that create significant operational pressure for system administrators. Daily troubleshooting and business adjustments require administrators to manually execute commands on each server in the data chain based on business documentation, resulting in low efficiency and high error rates.

Limitations of File-Centric Architecture: For years, the platform’s data transmission and processing services—whether between internal systems or with external users—have revolved exclusively around “files” as the data format. However, the mobile internet revolution has transformed news dissemination patterns. Message-driven architectures and API-based data interaction have become increasingly prevalent, while RSS feeds and data subscription models are now widely adopted. In contrast, the platform’s longstanding file-and-directory-based transmission model can no longer adequately meet business requirements, constraining its development.

Fragmented System Versions: As the central transmission hub, the platform interacts with numerous upstream and downstream systems, each with unique requirements: different transmission protocols (Socket or FTP), heterogeneous operating systems (Windows, Linux, Solaris), varied file landing specifications (date-based structures, language-based organization, multiple physical copies), and considerations for data verification and resumable transfers in international network transmissions. To accommodate these diverse needs, the platform has spawned multiple versions of its data transmission programs. While functionally similar, these versions differ in implementation details, making maintenance difficult and increasing the risk of errors during subsequent deployments.

Proliferation of Hardware and Network Complexity: Over the past decade, Xinhua’s rapid news business growth has spawned numerous technical systems with varying network architectures and personalized data interaction requirements. To address this, the data exchange platform has deployed node servers across six network zones: internal network, DMZ, external network, green zone interaction area, green zone application area, and private network, with one or more servers in each zone. Currently, the platform operates over 20 node servers handling data exchange tasks, plus more than 10 access switches across network domains. This hardware proliferation, combined with applications written primarily in C language and shell scripts, creates significant operational overhead.

Future Development Directions

Service-Oriented Architecture (SOA) with Unified Management:

The platform should adopt a Service-Oriented Architecture framework built on loosely coupled principles, providing distinct services for data ingestion, format conversion, transmission, callback, query, and retrieval. This approach enables cross-platform data exchange services with centralized, unified control and standardized management of the entire data lifecycle—from ingestion through processing and transmission.

Unified Storage Implementation: A major source of architectural complexity is that transmitted data resides in local file systems of various systems, necessitating deployment of transmission nodes across network domains to shuttle data back and forth. Introducing unified storage (such as NAS) addresses this

challenge. As an independent storage network separate from the server business network, it allows servers in different network domains to simultaneously access the same storage network, enabling data sharing. This significantly improves data access timeliness while drastically reducing inter-server network transfers and conserving network resources. Additionally, NAS provides fine-grained access control through configurable read, write, and execute permissions for different users, ensuring data security. Consequently, node servers are only required for network domains that cannot yet access unified storage, substantially reducing server deployment counts.

Distributed Deployment and Clustering: Leveraging unified storage for both content data and system application data enables convenient sharing between servers. The platform's computing resources can therefore be distributed according to service functions and implemented in cluster mode. This offers several advantages: First, functional distribution creates loosely coupled application modules with clearer business management logic and faster problem localization. Second, with shared databases and configuration files, servers can achieve "active-active" or even "multi-active" operation, significantly improving stability compared to traditional one-to-one cold standby. When one server fails or crashes, others in the cluster can immediately take over without business interruption. Furthermore, clustering provides the foundation for load balancing, which is crucial for high-traffic core business nodes.

Message-Driven Architecture and API Services: Building upon the traditional file-centric model, the platform should develop and prioritize a message-centric business model. Using mature message-oriented middleware, data interaction and service orchestration—both internally and with external systems—can be implemented through messaging. The aforementioned distributed deployment and server clustering represent prime examples of message-driven architecture. Additionally, frequently used functional modules such as format conversion, data distribution, and even data transmission should be encapsulated as services and exposed through APIs for consumption by relevant systems. From the perspective of integrated systems, this approach allows flexible data processing according to their specific needs while enabling the platform to focus on maintaining core functional modules and interface stability without accommodating every system's personalized requirements, resulting in clearer system boundaries.

Platform Refactoring with Java/J2EE: The platform should be refactored using Java and standard J2EE specifications to ensure smooth cross-platform deployment without operating system limitations. During refactoring, personalized functionalities should be implemented through comprehensive configuration files rather than hard-coded in the main program, with corresponding extension points reserved. This ensures that business adjustments require minimal modifications to the main program, focusing instead on configuration file updates. Such an approach maintains stable and unified application versions while facilitating packaging and deployment as agents on interface servers of relevant

systems.

Unified Monitoring with ELK Stack: Integrating the ELK (Elasticsearch, Logstash, Kibana) real-time log analysis platform enables more efficient and convenient business monitoring. FileBeat, a lightweight log collection tool installed as an agent on target servers, collects logs and forwards them to Logstash. Once aggregated and stored, Kibana provides a web interface for log analysis, helping administrators summarize, analyze, and search critical data. To maximize efficiency, all application logs must follow unified output standards.

System hardware monitoring should track metrics including disk usage and memory utilization, obtainable through OS message extraction and simple shell commands. For business monitoring, refactored modules must output logs in standardized formats with clear, concise fields related to business faults, enabling precise and rapid fault retrieval in the ELK platform. Since log information from all servers converges in one place, administrators can achieve centralized visibility and trace data flows across servers, quickly identifying and resolving issues.

References

- [1] Thomas Erl, Li Dong, Li Duo (trans.). SOA Architecture: Service and Microservice Analysis and Design [M]. Beijing: Mechanical Industry Press, 2017, (2): 10-26, 40-72.
- [2] Bruce Eckel, Hou Jie (trans.). Thinking in Java [M]. Beijing: Mechanical Industry Press, 2002 (2): 651-667.
- [3] Ni Wei. Distributed Message-Oriented Middleware Practice [M]. Beijing: Electronic Industry Press, 2018 (1): 23-42.
- [4] Bing Xin Wu Ying. Detailed Explanation of ELK Log Analysis System [EB/OL]. 2017-04-24, <https://blog.51cto.com/12854546/1918773>.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.