Public Cloud-Based News Service Platform Architecture Design (Postprint)

Authors: Yang Qi

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

Abstract

Against the backdrop of turbulent international developments, strengthening China' s media influence overseas is of particular importance. This paper optimizes the design of Xinhua News Agency' s news feed service architecture based on public cloud, comprehensively accelerating overseas content delivery speed and news dissemination effectiveness, improving overseas users' access experience, and enhancing Xinhua News Agency' s international communication capacity.

Full Text

Preamble

Architecture Design of a News Service Platform Based on Public Cloud

(Xinhua News Agency, Beijing 100803)

Abstract: As international dynamics grow increasingly complex, strengthening China' s media influence overseas has become particularly critical. This paper presents an optimized architecture for Xinhua News Agency' s feed service based on public cloud infrastructure, comprehensively accelerating overseas manuscript delivery and news dissemination effectiveness while improving access experience for international users and enhancing Xinhua's global communication capacity.

Keywords: news service; public cloud; platform architecture design; user access experience
CLC Number: TP237
Document Code: A
Article ID: 1671-0134(2021)03-109-03
DOI: 10.19483/j.cnki.11-4653/n.2021.03.030

Citation Format: Yang Qi. Architecture Design of a News Service Platform Based on Public Cloud [J]. China Media Technology, 2021(03): 109-110, 113.

Introduction

The current international landscape is turbulent, making it imperative to strengthen China' s media influence abroad. Meanwhile, the rapid development of cloud computing and internet technologies has forced traditional media to adapt to a fundamental shift in user behavior from print to screen. In today' s era, mastery of the internet means mastery of strategic initiative, and cloud computing serves as a critical foundation for internet development. Consequently, a news service platform architecture based on public cloud has emerged as the natural solution.

2.1 News Service Platform

The news service platform (also called the feed platform) is an internet-based technical system that provides news organizations with online feed services and information data push services, encompassing all Xinhua' s textual, photographic, graphic, video, new media, and multimedia content, along with historical archives and third-party products.

The platform' s primary business scenarios include: (1) online push, (2) online browsing, (3) user subscription management, (4) online assignment, (5) online product packaging and publishing, (6) user management, (7) online communication and notifications, (8) new media content display, (9) online search, (10) operational analysis, (11) third-party platforms, and (12) service capability monitoring. The platform generates approximately 40GB of manuscript data daily and retains three months of production data. Manuscript data is categorized into raw data and metadata. Raw manuscript data consists of XML files and attachments, where XML files describe the manuscripts and attachments primarily include images, videos, audio, HTML/CSS/JS, PDF, CDR, ZIP files, and other formats. Text content is stored in NoSQL and MySQL relational databases, while image entities reside on NAS storage with their metadata stored in MySQL. Other attachment metadata is similarly stored in MySQL, with physical files stored on NAS. Images require generation of medium and thumbnail versions for web display in addition to external distribution. Videos may require transcoding and keyframe extraction beyond external distribution.

The news service platform delivers data services to overseas institutional users through client synchronization, email, RSS, API interfaces, and FTP push, supporting formats including XinhuaML, CNML, NewsML, SimpleML, RSS, IPTC, and plain text.

2.2 Architecture Design

The news service platform employs a public cloud-based architecture, rapidly establishing a scalable, globally distributed cloud feed service with three core sites: Asia-Pacific (Hong Kong), North America (United States), and Europe (France) [Figure 1: see original paper], enabling rapid global manuscript distribution and providing overseas users with one-stop cloud feed services. The Beijing site is Xinhua' s proprietary facility, maintaining substantially identical manuscript data with the Asia-Pacific, North America, and Europe sites, though cached hot content varies based on regional user access patterns. User behavior data—including downloads, browsing, and comments—is transmitted back to the Beijing site.

2.2.1 Network Customization

The platform leverages public cloud network customization capabilities to define, plan, and manage networks according to deployment requirements, including custom subnet addressing and security group ACL rules. Virtual Private Clouds (VPCs) in public cloud enable construction of one or more isolated network environments, ensuring security while meeting autonomous network definition needs. Based on the platform's requirements, separate VPCs were established in Asia-Pacific, North America, and Europe. The application system is divided into publishing, application service, and infrastructure component layers, with three corresponding security groups implemented through security group technology to ensure logical business isolation, minimal privilege access, and security. The internet access zone utilizes public cloud DDoS protection and WAF services, while the internet egress employs dynamic BGP to guarantee high availability and low latency unaffected by carrier networks.

2.2.2 Manuscript Data Synchronization

Manuscript data synchronization relies on underlying transmission network infrastructure, yet the international internet environment is complex and unstable. To ensure stable synchronization, the platform uses public cloud dedicated line interface services to connect the Beijing site with the Asia-Pacific site (Hong Kong) through two redundant dedicated lines, while interconnecting the Asia-Pacific, Europe, and North America sites via public cloud cloud 专线 services.

The manuscript data synchronization process [Figure 2: see original paper] proceeds as follows: (1) Create a MongoDB database at the Hong Kong site to store manuscript files; (2) The Beijing master station' s feed system synchronizes manuscript data to the Hong Kong database; (3) The feed platforms in Hong Kong, France, and the United States all load and retrieve manuscripts by accessing the Hong Kong database.

2.2.3 High Reliability Design

The platform' s high reliability design encompasses two levels. First is the public cloud infrastructure' s native reliability: (1) Anti-affinity policies distribute instances across different hosts to minimize impact from node failures—elastic cloud servers within server groups follow anti-affinity strategies to maximize reliability; (2) Cloud server HA automatic recovery migrates virtual machines with HA attributes to healthy compute nodes when a node fails or a VM enters an error state, ensuring rapid recovery—enabling this feature during VM creation safeguards business continuity; (3) Data backup through VM data backup and management data backup ensures recoverability, allowing management nodes to automatically restore data to backup points when needed.

Second is the multi-site reliability design for the news service platform itself: through global load balancing, overseas users can access the nearest site more efficiently, and when a site fails, traffic can be quickly switched to alternative sites.

3. Application Effects and Benefits

Public cloud resources enable rapid deployment, dramatically reducing setup time compared to traditional physical servers and accelerating migration of all overseas feed users to the cloud. The public cloud-based news service platform achieves unified operations and maintenance across global sites with unified user services, improving service quality and responsiveness for feed subscribers. Centralized maintenance of multiple global sites not only significantly reduces operational costs but also enhances technical support for overseas marketing.

Leveraging the extensive global coverage of cloud nodes, the news service platform can flexibly select deployment locations based on user distribution, enabling users to access cloud feed services from the nearest point. The platform' s flexible scalability allows dynamic and efficient optimization of CPU, memory, storage, and network resources based on actual operational conditions, ensuring optimal performance at all times—particularly crucial during major coverage periods when important manuscripts generate surging user traffic. Public cloud services adopt a priority service and focused support principle, proactively optimizing and expanding resources to provide quality service during critical periods.

As public cloud services continue to optimize, user access to overseas sites shows significant improvement compared to accessing the Beijing site: average latency for web systems and manuscript FTP push has been reduced by 2-3 seconds, with speed improvements exceeding 60%. This has achieved excellent social benefits and enhanced the image and brand competitiveness of Xinhua' s feed services.

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