# Research Postprint on Cloud Computing Power News Big Data Platform

**Authors:** XU Zhiqiang, Zhang Shouxian, Li Manjiang

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

## Abstract

Different news big data companies exhibit varying data processing capabilities and specialized domains. This platform investigates how to leverage the respective strengths of each company to achieve a news big data platform that is more powerful in functionality, more comprehensive in data, more accurate in analysis, and faster in speed, in a cost-effective manner.

## Full Text

### Research on Cloud Computing Power News Big Data Platform

**Abstract:** News big data companies exhibit varying data processing capabilities and specialized domains. This study investigates how to leverage these diverse capabilities to deliver a more powerful, comprehensive, accurate, and efficient news big data platform at reduced cost. The platform can initiate data requests to major big data providers, obtaining optimally processed results that have undergone intelligent analysis, filtering, deduplication, localization, and targeted processing. This creates the most comprehensive, capable, professional, and timely public opinion service, offering multi-dimensional analyses—including latest news, rolling headlines, regional news, dissemination rankings, propagation paths, trend analysis, hot topics, public welfare hotspots, sentiment analysis, search trends, hot keywords, regional rankings, regional hotspots, and popular figures—across diverse data sources such as news websites, government portals, search engines, forums, microblogs, WeChat, and news apps.

**Keywords:** Cloud computing power; News; Big data; Cloud computing

**CLC number:** TP311
**Document code:** A
**Article ID:** 1671-0134(2019)09-116-02
**DOI:** 10.19483/j.cnki.11-4653/n.2019.09.034

**Authors:** Xu Zhiqiang[1], Zhang Shouxian[2], Li Manjiang[1]

Products from major technology companies such as Baidu, Tencent, Sina, NetEase, Sohu, and Toutiao provide news-related services including public opinion monitoring, trending topics, hot searches, briefings, headlines, and rankings. However, each company' s distinct data sources result in specialized strengths: Sina' s massive Weibo dataset makes its "Yuqingtong" product particularly authoritative for analyzing hot topics and influence on Weibo, while Tencent' s big data platform excels in analyzing vast amounts of news content from WeChat Official Accounts. A comprehensive news big data platform requires extensive data coverage across websites, microblogs, WeChat, apps, official accounts, and forums. Relying on a single data source yields insufficient data and incomplete, inaccurate analysis results. Greater data comprehensiveness leads to more accurate, timely, reliable, and objective results. This platform research explores how to integrate capabilities from multiple big data companies through a unified platform, combining integration, segmented procurement, and customized crawling to significantly reduce procurement costs for news organizations and government agencies while delivering superior public opinion services.

## 1. What is Cloud Computing Power?

Big data analysis requires substantial computational resources to process extremely large volumes and deliver results within short timeframes. The term "computing power" primarily refers to computational capacity—for instance, the computing power of Bitcoin mining machines (hash rate) measures the processing capability of the Bitcoin network, representing the speed at which a computer (CPU) can compute hash function outputs. In this paper, "cloud computing power" adapts this concept to describe the integration of independent cloud computing capabilities from multiple companies, forming a unified, cloud-based data processing capability.

## 2. Platform Research

This research focuses on a comprehensive platform built upon the computing capabilities of various news big data companies. By intelligently leveraging these providers' computing power, the platform enables news organizations, government agencies, and other entities requiring public opinion services to use a single cloud-based interface. Users can initiate data requests to multiple big data companies simultaneously, obtaining optimally processed results that have undergone intelligent analysis, filtering, deduplication, localization, and targeted processing. This creates the most comprehensive, capable, professional, and timely public opinion service, offering multi-dimensional analyses across diverse data sources.

### 2.1 Feasibility

**Technical Feasibility:** The platform relies on the processing capabilities of other companies. For a big data company to be integrated into the platform, it must provide open APIs or data push methods; otherwise, the platform would need to crawl result web pages and store them independently.

**Commercial Feasibility:** Purchasing big data services from various companies typically does not restrict multiple displays to different commercial users through a single platform in licensing agreements. After procuring data services from multiple providers, the platform can comprehensively organize and present integrated results to different target customers through separate accounts. Through multiple sales, the costs of purchasing data services can be amortized, enabling high-quality news public opinion services at low prices.

### 2.2 Implementation Plan

Platform implementation encompasses more than data integration, involving several key aspects:

(1) **Existing Services and Integration of Each News Big Data Company:** Certain existing services from each company can be directly presented to users without additional processing, maintaining their authority and meeting user requirements.

(2) **Processing Capability Integration:** Based on each company' s interface methods, the platform completes integration of their data services, utilizing APIs, web crawling, and other methods to obtain data.

(3) **Data Cleaning, Deduplication, and Metadata Standardization:** Results from big data companies must be merged into consistent data. Different companies employ inconsistent data definitions—for example, in basic information, date formats, and scoring ranges. The platform must unify formats, remove duplicate data, and cleanse it into consistent, valid datasets.

(4) **Integration of Processed Results from Various Parties:** This involves weighted synthesis, localization, and targeted processing. When presenting data from multiple companies, the platform integrates it through weighting. Based on user needs, irrelevant regional data is removed, retaining only data pertinent to the user' s region of interest.

(5) **Comprehensive Scheduling:** When users request a particular service, the platform backend can simultaneously initiate service requests to multiple big data companies as needed, then integrate and present the returned results.

Additional functionalities include personalized local data crawling, customized news public opinion services, multi-tenant management, and data isolation for different user organizations.

*Funding: This research is supported by the Weifang Science and Technology Development Plan Project (Project No. 2019ZJ1162).*

**Platform Function Example:** To demonstrate a news item' s dissemination effect, the platform must incorporate propagation data from multiple channels —including newspapers, websites, WeChat, microblogs, apps, forums, and social networks—to comprehensively showcase impact, breakout points, and timelines. Audience distribution and timeliness vary across channels, with news propagation exhibiting distinct periods of breakout, diffusion, decline, and termination. After crawling results from various platforms, weighted integration creates a relatively complete and objective timeline curve while preserving individual channel timelines for reference. Metrics such as click counts, comment numbers, and audience profiles from each channel are integrated and presented through charts, tables, and graphs.

In terms of investment, the cloud computing power news big data platform requires significantly lower capital than individual big data companies' analysis platforms. Demands for data storage and computational capacity are substantially reduced, yet the integrated results are not inferior and even enhance services compared to single-company offerings. This model depends on a sufficient number of end users to amortize service fees from various big data providers. Additionally, platform development and maintenance entail considerable workload.

**References:** [1] Sun Qihu. Research on news media production and dissemination strategies in the big data era [J]. *Journal of Shandong Agricultural and Engineering University*, 2019(2).
[2] Sun Yan, Li Kuishang, Sun Jiancai. Using collaborative space to complete cross-newspaper coverage of major events [J]. *China Media Technology*, 2018(2).
[3] Li Zhiqiang. Experience in news field reporting command systems [J]. *China Media Technology*, 2017(8).

(Author affiliations: 1. Weifang Beida Jade Bird Huaguang Phototypesetting Co., Ltd.; 2. Peninsula Metropolitan Daily)

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