

Design and Application of ElasticSearch Cluster Automated Operation and Maintenance Management System Based on Tianjing DevOps Platform

Zhe Niu, Jie Fu*

*Gansu Meteorological Information and Technical Equipment Support Center, Lanzhou, Gansu,
China*

**Corresponding author*

Keywords: Tianjing; ElasticSearch; Automated Operation and Maintenance; Automated Patrol Detection; DevOps

Abstract: As the most important monitoring operation and maintenance platform in meteorological operation, Tianjing ensures the stable operation of meteorological core operation system. However, with the huge growth of meteorological data, the rapid growth of the indexes of the Elastic Search (ES) cluster of Tianjing has a great impact on the stability of the cluster. Based on the Tianjing DevOps platform, an intelligent operation and maintenance management system for ES database is researched and designed, which can automatically realize the production line work of fault alarm, detection, treatment and prevention, which can improve the efficiency of operation and maintenance and the utilization rate of resources, and ensure the stable operation of ES database cluster.

1. Introduction

Meteorological information is the “center” of meteorological operation. It is an important way to realize the high-quality development of meteorological undertaking by promoting the transformation of traditional meteorological operation to the new operation mode of intelligent digital meteorology with informatization [1]. The key point of meteorological informationization is to strengthen scientific management and efficient application of massive meteorological data [2], so as to not only meet the requirements of stable acquisition and timing of data resources of meteorological products and applications, but also ensure the long-term archiving security of basic data and key products. Therefore, the database maintenance strategy and efficient management scheme are extremely important in the big data cloud platform [3].

Through deep processing and integration of meteorological information resources, the provincial integrated meteorological operation real-time monitoring system provides users with the whole-workflow cloud+end computing service of “integration of data and calculation”. The calculation results will be stored in ElasticSearch (ES) database cluster. The system obtains corresponding monitoring data from ES and displays them, so as to realize the functions of monitoring of the “whole-workflow, all elements, whole process ”[4]. As the main monitoring information data source, the massive index storage causes a certain burden on the ES. Meanwhile, in order to ensure the normal

operation of the whole workflow monitoring, the running conditions of relevant nodes and processes need to be monitored in real time. Therefore, optimizing the drive space of the ES, and automatically process the daily operation and maintenance contents of the ES database, can improve the work efficiency of the operation and maintenance personnel [5]. At the same time, the continuous development of the DevOps automated operation and maintenance platform of the provincial meteorological integrated operation real-time monitoring system can give full play to the resource advantages and facilitate management and upgrading. And the system can be applied to actual operation and is easy to transplant and be promoted to the whole province.

In addition, automated operation and maintenance has become the development trend in the future. Automating a large number of repetitive tasks in IT operation and maintenance, and transforming the past manual execution into automated operation, which is a sublimation of monitoring and maintenance work, also an improvement of the information management process [6].

2. DevOps Automated Operation and Maintenance Platform

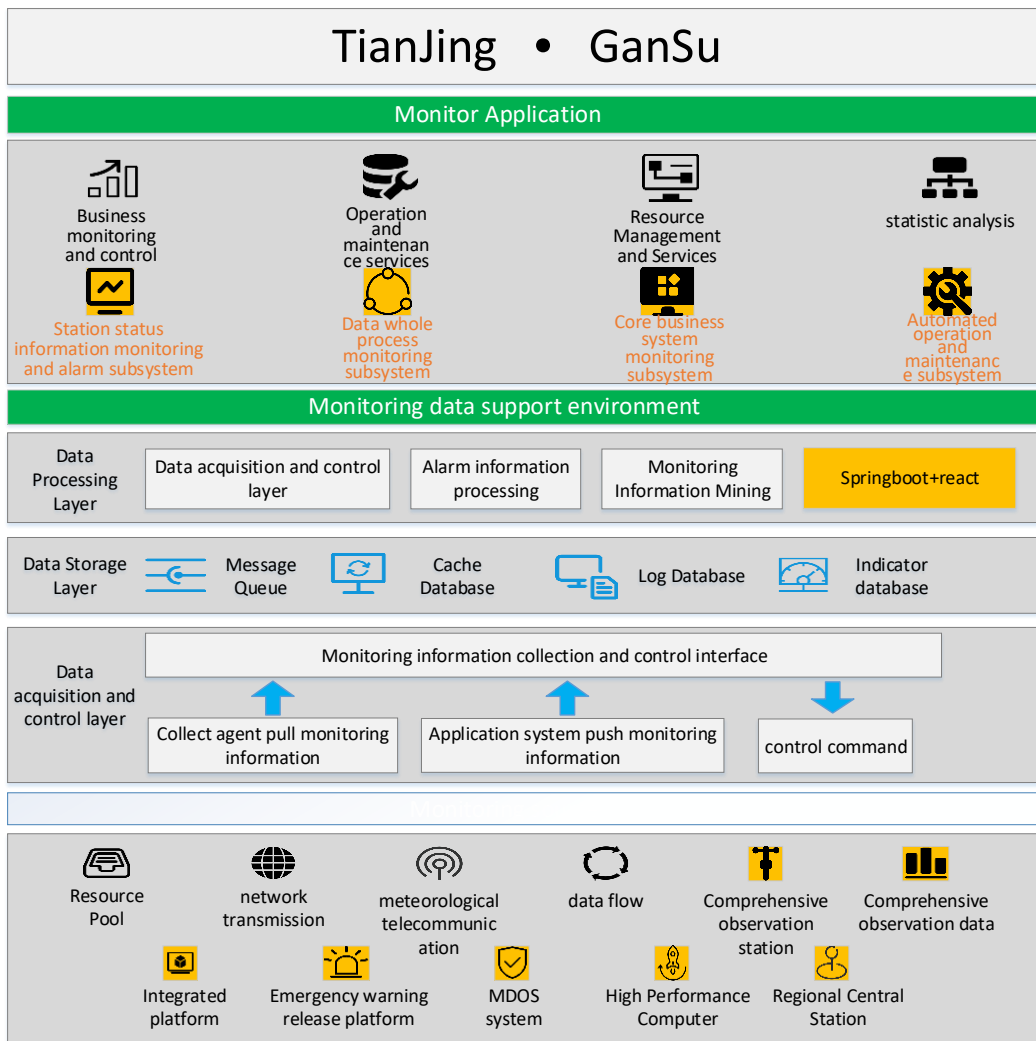


Figure 1: System architecture

DevOps concept was proposed in 2009. It emphasizes in-depth communication between software

developers (Dev) and operation and maintenance technicians (Ops), realizes production line style management of project development process, provides unified and standardized integration and service work in the process from development, test to operation and maintenance, and is widely used in automated operation and maintenance management [7]. The real-time monitoring system for Gansu Meteorological Comprehensive Business is based on the DevOps concept and provides an automated operation and maintenance platform. It adopts SOA (service-oriented architecture), J2EE architecture, and Spring Boot components for hierarchical and componentized application software construction. Business personnel can develop standardized scripts and choreograph scenarios related to operation and maintenance in actual business on the platform, achieving an effective combination of resource monitoring and automated collaboration. The system architecture consists of the following parts (as shown in Figure. 1):

(1) Monitoring source: The "Tianjing" provincial-level universal version platform currently supports data collection for resource pooling, network transmission, meteorological communication, and the entire data process. On this basis, this project has added 15 new data collection items, including the four main subsystem operation indicators of the Tianqing system.

(2) Data acquisition and control layer: A unified acquisition and control platform is established based on data acquisition business. The platform adopts a distributed resource acquisition and control system to achieve unified acquisition and control of managed resources, and supports third-party system integration and management. The unified acquisition and control platform provides a unified channel for communication between various operation and maintenance tools and managed equipment resources, and allows each operation and maintenance tool to freely expand its acquisition and control capabilities through module and plugin technology, without paying attention to the underlying communication and scheduling technology. Instead, it only needs to write acquisition and control scripts according to the agreed specifications of the acquisition and control module, organize them into strategies, and distribute them to the corresponding agents, and process the resulting data to complete machine data collection Configuration change release and resource operation control.

(3) Data storage layer: The data collected by the data acquisition and control layer adopts the stream processing engine technology, which is based on Redis+Gemfire Distributed cache technology. Through clustering, load balancing and other methods, it can improve the system access speed, enhance the system disaster tolerance, reduce the pressure on the data server, and achieve dynamic scalability of the cache. Implement new collection data (model data, indicator/alarm data, and file data) to be written into Elasticsearch and MySQL databases respectively.

(4) Data processing layer: the localization application modules are uniformly deployed based on the mass innovation platform. Data processing methods, including stream based index calculation, alarm information processing, monitoring information mining and other mechanisms, adopt the same standards as the national sky mirror or special standards of provinces and cities.

(5) Monitoring application: In addition to monitoring applications such as business data process monitoring, operation and maintenance service monitoring, resource management and service monitoring, and core business system monitoring on the "Tianjing" provincial-level universal version platform, based on newly collected data types, develop Tianqing data full process centralized monitoring and alarm display system application services.

The automated operation and maintenance platform (as shown in Figure. 2) collects the configuration information issued by the management gateway using the method of deploying agent agents to perform tasks on the target servers [8]. Its core concepts include hosts, operations, files, arrangement, and assignments. Host is the target object of the task, and the operation is an actual processing action for the target. The arrangement is the tasks entity, which is the operation and maintenance scene that the user needs to implement. An arrangement can be a single operation or a combination of multiple operations. Each run of the operation and maintenance task is called an

assignment, and the specific execution process of the whole task is recorded in the assignment. To develop scheduled tasks on the automated operation and maintenance platform, the development environment needs to be prepared according to the expected targets. The target server needs to be the node managed by the management gateway. Therefore, it is necessary to add nodes in the acquisition and control platform module in advance, deploy the agent in the mode of local agent, open the authority of local operation and local monitoring to it, and set the agent to be always in the running state after installation. After setting the environment, add new operations in the platform. After defining the basic parameters, add the script content for testing. After the adding operation is completed, the newly added operation is added to the arrangement according to the task flow to complete the addition of the operation and maintenance scene.

The automated operation and maintenance platform supports a variety of operation and maintenance scenes. It can realize patrol detection management of basic resources such as servers, key processes or operating systems, and automatically complete assignment of running batch work such as daily patch upgrading and regular cleaning of logs. It is widely used in the provincial comprehensive meteorological operation monitoring system [9].

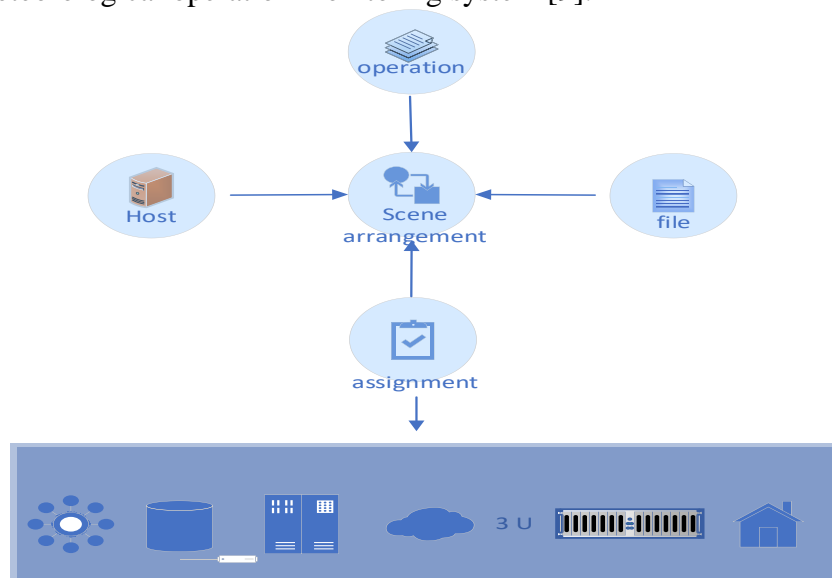


Figure 2: Panoramic view of automated operation and maintenance platform

3. ES Intelligent Automated Operation and Maintenance System

3.1. Operation Flow

Firstly, deploy a collection agent on the data source server to collect real-time meteorological business data, basic resources, network, computer room, and other monitoring data; The collection agent sends data to the access gateway by calling the data aggregation access interface; The gateway divides the data into indicator information and event information, and sends them to different queues through Kafka. Among them, indicators include timeliness indicators or quantity indicators, while events refer to specific events such as alarms. All events are stored in the ElasticSearch database, and then the event extraction indicators and threshold values are calculated through spark streaming and stored in the Cassandra indicator database. The data batch analysis module stores the analysis result data into the database through Spark SQL, and finally provides a unified restful API interface to provide external monitoring data services.

3.2. System Design

Meteorological integrated operation monitoring system shall process tens of millions of data according to operation requirements. Therefore, Elastic Search (ES) distributed database is adopted for storage to provide query optimization. Massive calculation results will be stored in ES cluster. The monitoring system shall acquire corresponding monitoring data from ES and display them. In the actual operation, because the index data volume is too large, it will cause certain burden to the cluster, causing abnormal health status of the cluster, thus affecting the normal use of the meteorological integrated operation system [10]. Therefore, it is necessary to develop a set of ES database intelligent automated operation and maintenance system to realize the whole production line operation and maintenance work of fault alarm, detection, treatment and prevention.

The design of ES intelligent automated operation and maintenance system mainly includes the following four aspects:(1) Monitoring the disk space of ES database, and alarming if occupied more than 80%; (2) Writing automation script, if the disk space exceeds the threshold value, automatically closing the unused index and releasing space; (3) Realizing the function of scanning node status regularly for ES database nodes; (4) If a node status is abnormal, the failed node will be identified and the process is automatically started.

Firstly, monitoring the disk space of the ES database, and adding operations in the automated operation and maintenance module. If the disk space of the server is greater than the set threshold, the shell script deployed on the server will be executed, and the unused index will closed to reduce the occupying space. Clean up the space in advance before the cluster status is abnormal to reduce the possibility of fault occurring.

The health monitoring of cluster status adopts the API tool of ES, to visit `http://*/cat/health`. As shown in Fig. 3, the health status of cluster can be obtained.

```
timestamp cluster status node.total node.data shards pri relo init unassign pending_tasks max_task_wait_time active_shards_percent
13:56:48 elasticsearch5 green 9 9 6534 3252 0 0 0 0 - 100.0%
```

Figure 3: Health Status

If the node status is green, you will be prompted that the status is healthy. If it is yellow, you shall access `http://*/cat/nodes`. As shown in Fig. 4 (ip is hidden), normally there should be 9 nodes. The missing node ip is obtained through a script, and then the alarm is given and the process of the corresponding node is started.

```
97 62 72.04 63.97 56.74 mdi * 5N2Tr1y
96 4 1.60 1.65 1.80 mdi - VFc3euP
97 49 48.10 33.91 28.76 mdi - q5JI2t9
97 1 0.42 0.78 0.92 mdi - 6-eZ-ib
100 3 6.61 7.56 11.29 mdi - nuwNsYo
97 3 2.57 2.67 2.68 mdi - RABcEXO
98 4 2.28 2.16 1.94 mdi - VtFkZev
94 3 0.92 1.14 1.29 mdi - gkCe-7C
98 56 74.55 61.03 48.44 mdi - -2EiEFD
```

Figure 4: Node Status

4. Summary

In this paper, the daily operation and maintenance contents of the meteorological integrated operation monitoring system ES database are sorted out. The timing monitoring script and the automated operation and maintenance strategy are set based on the DevOps automated operation and

maintenance platform, so the functions of automatically pulling up the processed and deleting invalid indexes are realized, the operation and maintenance process of the ES database is optimized, and the resource advantage of the platform is brought into full play. Meanwhile, the working efficiency of maintenance personnel is greatly improved. Therefore, this paper provides a solid foundation for guaranteeing the weather data transmission quality.

References

- [1] Zhang P. & Sun X. Q. (2023). *Research on the Construction and Operation of Quality Management System of Metrological Instrument Metrological Verification Institute*. *Meteorological, Hydrological and Marine Instruments* (01), 44-46. doi:10.19441/j.cnki.issn1006-009x.2023.01.005.
- [2] Zhang H. Z., Zhang Z. Q., Li J. & He X. M. (2016). *Open Special Report on Meteorological Data*. *China Informatization* (07), 72-75.
- [3] Shen W. H. (2015). *Discussion on the Future Infrastructure of Meteorological Operational Information System-The Role of Cloud Computing and Big Data in Meteorological Informatization*. *Advances in Meteorological Science and Technology* (03), 64-66.
- [4] Zhang E. H., Yin H. Y. & Li G. J. (2019). *Research on Retrieval Technology of Meteorological Data Based on Elasticsearch*. *Computer Technology and Development* (11), 154-158.
- [5] Fan L. (2020). *Elasticsearch Massive Data Storage Query Optimization*. *Industrial Control Computer* (10), 85-87.
- [6] Zheng P. F. (2020). *Application of Intelligent Operation and Maintenance Based on Zabbix in Air Traffic Control Information System*. *China New Telecommunications* (22), 13-14.
- [7] Peng Y., Lai L. J., Wang Q. & Guo X. F. (2023). *Design and Implementation of a Lightweight DevOps platform for Edge Cloud*. *Modern Information Technology* (05), 85-89. doi: 10.19850/j.cnki.2096-4706.2023.05.020.
- [8] Wang F., Zhang B. S. and Xu W. (2020) *Preliminary Study on Localized Development of Real-time Monitoring System for Meteorological Integrated Operations*. *Communication & Information Technology*, No. 243 (01): 39-40+74.
- [9] Zhang C. L., He X. T., Xie Y. H. & Tian J. (2022). *Research and Application of Key Technologies of Cloud Platform in Meteorological Information Infrastructure*. *Journal of Chengdu University of Information Technology* (05), 551-557. doi:10.16836/j.cnki.jcuit.2022.05.011.
- [10] Wang T. N., Li X. Q., Xu X. Q., Huang Y. H., Wang Y. P. et al. (2021) *Application of Automatic Warning Based on Cloud Platform of Meteorological Big Data*. *Ningxia Engineering Technology* 20. 04: 352-356.