ISSN: 1813-4890

# Fault Information of Oil and Gas Environmental Protection Equipment based on Hadoop Analysis Technology and Big Data Platform Design

Jian Zhang*, Xinyu Zhao

College of Computer Science, Yangtze University, Jingzhou, 434023, China

## Abstract

**Based on the application research of Hadoop architecture, combined with the fault information characteristics and diagnosis methods of oil and gas environmental protection equipment, and the analysis of existing problems in information analysis, combined with the comprehensive consideration of equipment fault information analysis and demand analysis, this paper proposes a big data architecture of oil and gas environmental protection equipment fault information analysis platform based on Hadoop, and discusses the fault information analysis technology.On this basis, the fault information is mined and analyzed by using Map Reduce-based architecture. Practice shows that the platform realizes the remote transmission, backup and extraction of multi-machine monitoring data information, the establishment of relational database and fault diagnosis, and has certain practical significance and practical application value.**

## Keywords

**Hadoop; Equipment Failure; Information Analysis; Big Data Platform.**

## 1. Introduction

In recent years, the fault diagnosis technology for oil and gas environmental protection equipment has made great progress in different aspects, such as signal acquisition and sensing technology, fault mechanism and symptom connection, signal processing and feature extraction, and intelligent decision-making and diagnosis.Key technologies of reliability, safety and maintainability of major products and facilities in oil and gas field are also listed as important research directions of the industry[1-2].At present, with the development of fault diagnosis technology, many intelligent diagnosis technologies such as expert system, neural network, Bayesian theory, fuzzy logic, rough set theory and genetic algorithm have emerged in the field of complex mechanical equipment. Through the analysis of these theories in academic circles[3], firstly, it is considered that the main cause of equipment failure and failure is the wear and vibration of mechanical equipment.Mechanical equipment failure caused by wear and vibration is accompanied by a great deal of capacity loss. Secondly, combining the advantages of wavelet packet decomposition and empirical mode decomposition in analyzing dynamic signals, the specificity of feature evaluation method in selecting sensitive features, and the strong classification ability of radial basis function neural network[4],An intelligent fault diagnosis model based on feature evaluation and neural network is introduced. Thirdly, aiming at the deficiency of fuzzy C- means (FCM) clustering algorithm[5], which assumes that each dimension feature is the same as each sample's contribution to clustering, and the number of clusters needs to be preset, three-layer feedforward neural network, point density function algorithm and clustering effectiveness index are used to improve it.A new hybrid clustering algorithm is introduced.

To sum up, at present, the fault diagnosis in the field of oil and gas environmental protection equipment is mainly based on traditional technologies such as signal processing and statistical

ISSN: 1813-4890

analysis, and the demand for fault diagnosis of large-scale complex systems still has great limitations[6]. Under the background of big data nowadays, the safety monitoring and fault diagnosis of oil and gas environmental protection equipment are facing the double innovation of theory and practice technology[7].

The analysis and processing method and system application of big data represented by deep learning simulates the thinking and learning process of human brain. By constructing a deep learning model, the complex mapping relationship between features and fault categories can be fitted, and the rich intrinsic information of big data can be depicted and the accurate identification of equipment faults can be realized[8].The determination of the remaining life of oil and gas environmental protection equipment is directly related to the safe operation of the equipment. From the big data of oil and gas environmental protection equipment, which contains a great deal of information about the safety state of the equipment and the performance degradation caused by faults, explore the evolution mechanism of the performance degradation of the equipment and parts, and establish a deep network[8]. Extract the characteristic index set related to characteristic structure life prediction from the massive monitoring data of independent oil and gas environmental protection equipment, quantitatively describe the fault change forms and trends, fully explore the state fault information hidden in the equipment big data, and provide strong data support for the life trend evaluation and prediction research of oil and gas environmental protection equipment.Promote the rapid development of related theories of life trend prediction of oil and gas environmental protection devices[9].

Based on the health monitoring and fault diagnosis system of oil and gas environmental protection equipment, this paper builds a big data platform to realize the remote transmission, backup and extraction of multi-machine monitoring data information, the establishment of relational database and fault diagnosis. It has strong practical significance and practical application value.

## 2. Design of Fault Information Analysis Platform for Oil and Gas Environmental Protection Equipment based on Hadoop

Combining with big data technology, based on the above-mentioned analysis of fault diagnosis characteristics, fault information characteristics and problems existing in information analysis of oil and gas environmental protection equipment, this paper proposes a Hadoop-based fault information analysis platform architecture of oil and gas environmental protection equipment, and the system architecture diagram is shown in Figure 1. The architecture designed in this paper combines the specific requirements of fault diagnosis and fault information analysis of oil and gas environmental protection equipment, and designs a fault information analysis platform of oil and gas environmental protection equipment based on Hadoop. On this basis, the fault information is mined and analyzed by using Map Reduce architecture.

The bottom layer of the whole architecture is the data layer based on Hadoop, which uses distributed file system HDFS and distributed database HBase to store fault information[10]. The support layer mainly realizes the processing of fault information by studying the key technologies of fault information analysis, and realizes the parallel mining analysis of fault information based on Map Reduce parallel framework. Using the data mining algorithm of clustering classification to realize the clustering and classification of fault types, so as to achieve the purpose of intelligent fault diagnosis; Its function layer and target service layer use some open user interfaces and visual charts to help users make decisions. The management of the platform monitors all levels of the platform to ensure the stable operation of the platform. All levels cooperate with each other to analyze and study the fault information of hobbing machine.

**ISSN: 1813-4890**

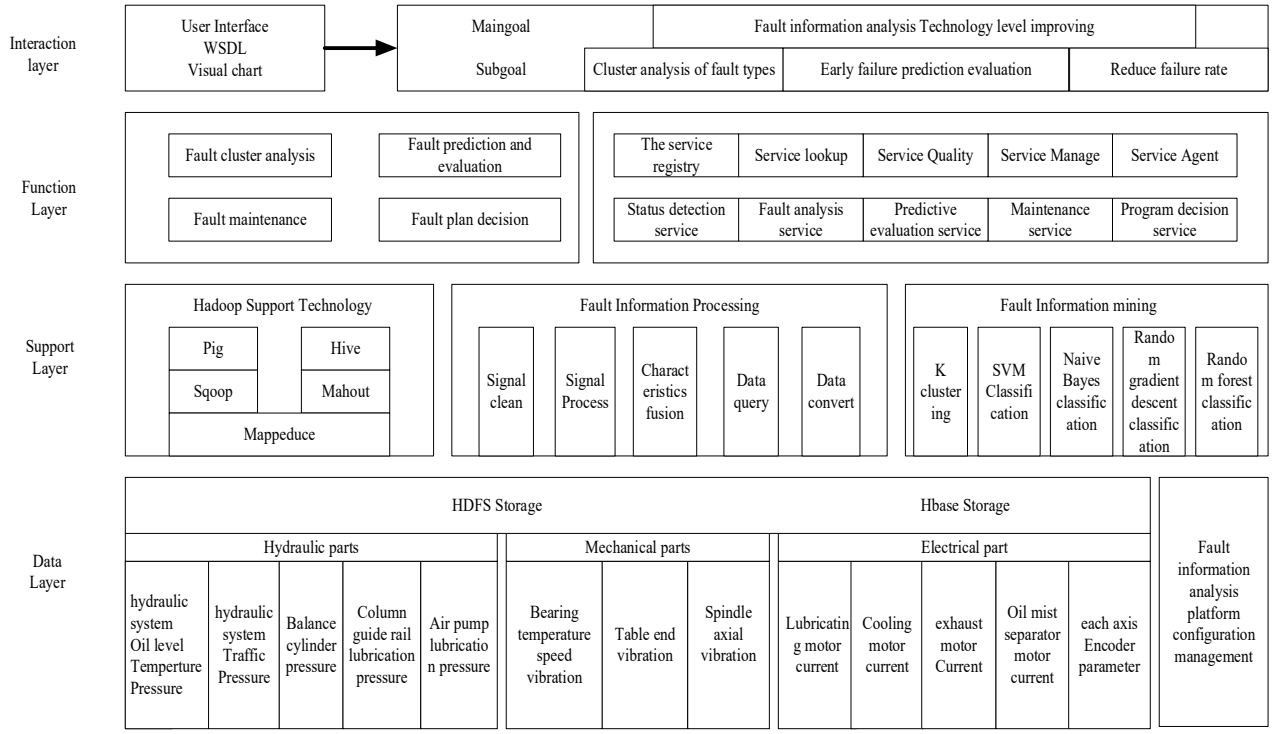


**Figure 1.** Architecture of fault information analysis platform for oil and gas environmental protection equipment based on Hadoop

## 3. Fault Information Storage Technology of Oil and Gas Environmental Protection Equipment based on HDFS and HBase

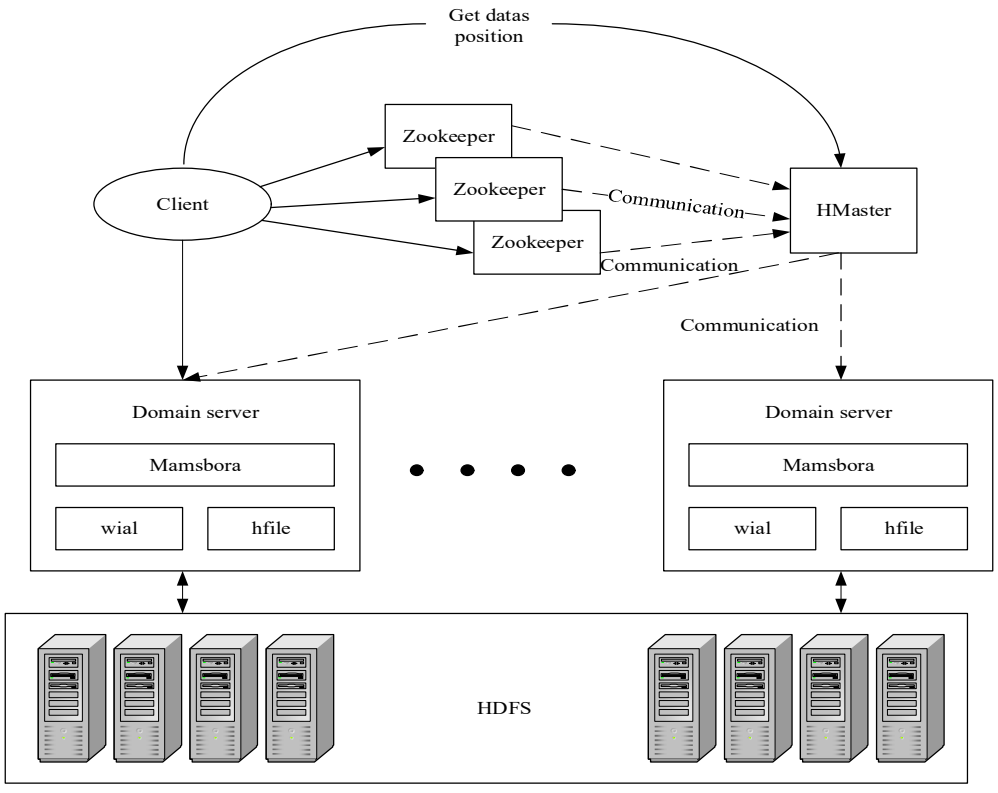


**Figure 2.** Fault information of oil and gas environmental protection equipment based on HDFS and Hbase Parallel storage process

ISSN: 1813-4890

HDFS distributed file system is one of the core technologies of Hadoop, which is used to provide data access services for massive network data processing platforms. Its design idea is based on Google file system. HDFS architecture is a master-slave architecture, including Name Node and Second Name Node. And Data Node. The Name Node is the master node, which is responsible for maintaining the namespace of the file system and managing the metadata of all directories and files. Second Name Node Is the standby node of Name Node, optional, and Data Node is the slave node, which maintains the management of its stored data blocks. HDFS files are stored in blocks, and files.

Divided into multiple fixed-size data blocks, the default is 64MB, and the data blocks are stored in different data nodes. At the same time, according to the number of copies configured by HDFS file system (the default is 3), the same data blocks are stored on multiple data nodes. HBase Is a distributed, versioned, column-oriented, multi-dimensional storage system, which is Open source implementation of Google Table architecture. Similar to HDFS, HBase implements master and slave (HMaster/ domain server) architecture. HBase data management is implemented by distributed domain servers, and domain servers are managed by HBase master server (HMaster).

HDFS can be used to store massive amounts of data that are basically accessed sequentially, while HBase the main advantage of is fast and random access to data. Combining HDFS and HBase can be used for efficient data storage.

## 4. Parallel Analysis Technology of Fault Information of Oil and Gas Environmental Protection Equipment based on Map Reduce

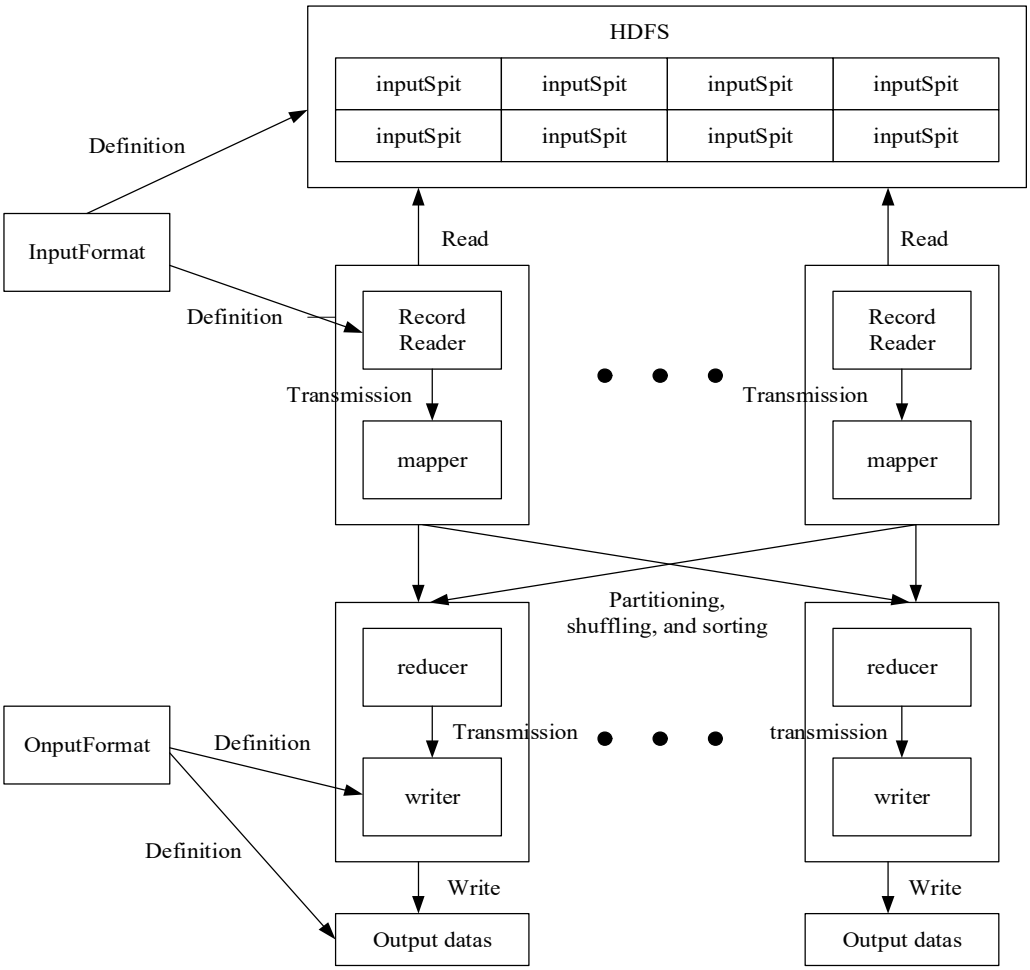


**Figure 3.** Parallel data processing flow based on Map Reduce

ISSN: 1813-4890

Map Reduce was first proposed by Google and actually used as a programming model for its network search service. It uses distributed parallel computing to process massive data of TB level or even PB level. Map Reduce distributed computing framework mainly includes map. And Reduce functions, users only need to write corresponding maps according to their own needs. And Reduce functions, that is, programs can be run in Hadoop platform to obtain data analysis results. The Map Reduce framework of Hadoop platform adopts master-slave architecture, which consists of one.

The Job Tracker master node and multiple Task Tracker slave nodes. Job Tracker The module is responsible for the scheduling of Map Reduce jobs, including the order of job execution submitted by users, the allocation and execution of Map tasks and Reduce tasks, and the execution of speculative tasks. In the slave node, the Task Tracker module is responsible for handling the tasks assigned by the master node, including the execution of Map tasks, Reduce tasks and speculative tasks. Map Reduce.

The framework operation flow is shown in Figure 3.

## 5. Conclusion

This paper starts with the analysis of fault diagnosis and fault information characteristics of oil and gas environmental protection equipment, expounds the existing problems of fault analysis at present, and then studies the preprocessing technology of fault information of oil and gas environmental protection equipment, and at the same time, analyzes the requirements of fault information platform, and puts forward the overall framework of fault information platform of oil and gas environmental protection equipment. Then, the detailed design scheme of using distributed Hadoop to build fault information platform is put forward, focusing on the research of distributed storage HDFS technology of equipment fault information and Map Reduce technology of fault information parallel analysis framework.

## Acknowledgments

Supported by Science and Technology Research Project of Education Department of Hubei Province - Guiding Project No. B2021041.

Supported by Humanities and Social Sciences Research Project of Education Department of Hubei Province No.20D038.

Supported by 2021 Science and Technology Research Project of JingZhou Science and Technology Bureau No. 2021CC28-16.

Supported by Industry-University Cooperation Collaborative Education Project of Ministry of Education in 2020 No. 202002077032.

Supported by Industry-University Cooperation Collaborative Education Project of Ministry of Education in 2021 No. 202102077032.

## References

[1] ZHANG Xiaofan, research and application on monitoring and diagnosistic methods of diesel typical fault based-on vibration impact features [D].Beijing university of chemcial technology,2018.

[2] Wang Z , Yang N , Li N , et al. A new fault diagnosis method based on adaptive spectrum mode extraction[J]. Structural Health Monitoring, 2021:147592172098694.

[3] Wang Z , Zhou J , Du W , et al. Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution[J]. Mechanical Systems and Signal Processing, 2022, 162(5-8):108018.

ISSN: 1813-4890

[4] X Wang, Mao D , X Li. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network[J]. Measurement, 2021, 173(6):108518.

[5] Levent E , Turker I , Serkan K . A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier[J]. Journal of Signal Processing Systems, 2018.

[6] WANG Yishou, QING Xinlin. Progress on study of structural health monitoring technology for composite joints[J]. Acta Materiae Compositae Sinica, 2016, 33(1): 1-16.

[7] DENG Dapeng, NIAO Xiaomin, ZHANG Jiancheng. Design and Simulation of signal conditioning Circuit for Interferometric Optical Fiber disturbance Sensor [J]. Modern Electronics Technique, 2011, 34(04): 28-32.

[8] Huang, Yu-Jiao; Wang, Xiao-Yan; Long, Hai-Xia. Synthesization of high-capacity auto-associative memories using complex-valued neural networ[J]. Chinese Physics B, 2016, 25(12): 12-16.

[9] Zhang Hai-yan, SUN xiuli, Fan shixuan. Lamb Wave Arrival Time Extraction Using Hilbert-Huang Transform for Improved Tomography Image[J]. Journal of Donghua University(English Edition), 2009, (05): 503-508.

[10] Liu, Mingzhou; Yang, Jiangxin; Cao, Yanpeng. A new method for arrival time determination of impact signal based on HHT and AIC [J]. Mechanical Systems And Signal Processing, 2017, 86. 177-187.