

A Fast Fault Diagnosis Method for Electric Energy Metering Device Based on Spark Streaming

Zeyuan Duan^a, Dewen Wang^b

School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

^aemail:121322016@qq.com, ^bemail:wdeven@gmail.com

Keywords: Spark Streaming; Stream Processing; Fault Diagnosis; Electric Energy Metering Device

Abstract. We present a fast fault diagnosis framework based on Spark Streaming for real-time online monitoring and fault diagnosis of electric energy metering device. In this framework, data collected by the power consumption information collection system is used as experimental input data, which is simulated as a real-time data stream. The BP neural network is applied to the fault diagnosis of metering device on the Spark Streaming cluster. The experimental results show that the processing speed of this stream processing cluster mode is higher than that of the traditional stand-alone mode, which can satisfy the requirement of fast fault diagnosis of massive data in smart grid.

Introduction

With the rapid development of smart grid, computer technology, sensor technology, etc., the data generated by the electric power industry shows explosive growth. With the application of the power consumption information collection system, a lot of business functions based on the collected data are realized [1]. At the same time, the system data collection is growing at an astonishing rate, which shows the characteristics of large data quantity and high data production rate. The data has the typical feature: real-time, volatile, disordered and so on [2]. However, due to the failures of collection terminals and metering devices, which affect the efficiency and authenticity of data collection, on-site maintenance has a lot of pressure [3] [4]. Therefore, real-time on-line monitoring and fault diagnosis of the electric energy metering device is of great significance to the development of smart grid [5] [6].

Spark Streaming is a real-time streaming computing framework built on Spark. It supports a variety of data input sources and data storage locations. The most striking feature of Spark Streaming is its high throughput and strong fault tolerance. In addition, Spark Streaming can be perfectly combined with other sub-framework of Spark, and hence is very suitable for real-time fault diagnosis of the metering device. This paper proposes a fault diagnosis framework for electric energy metering devices based on Spark Streaming and establishes a stream computing cluster. BP neural network algorithm [7] is deployed on the cluster, used for fault diagnosis of metering devices.

Stream processing framework: Spark Streaming

Spark Streaming is one of the Spark's sub-frameworks, which is mainly for distributed stream processing. It has a lot of advantages: it (1) can run on a large number of nodes, and reach the level of second delay, (2) uses memory-based Spark as an execution engine with high throughput and strong fault tolerance, (3) is able to integrate Spark's batch and interactive queries, and (4) provides interfaces for implementation of complex algorithms. Spark Streaming works by decomposing the incoming data stream into a series of smaller batch jobs. In fact, the input data is batch processed at a certain interval. When the batch time interval is shortened to the second level or millisecond level, it can be used to deal with real-time data stream. Each batch of decomposed data is then converted

into RDD in Spark. Finally, all the transformations of DStream in Spark Streaming change into the Transformations of RDD in Spark. The intermediate results will be temporarily stored in memory. Its working principle is shown in Figure 1.

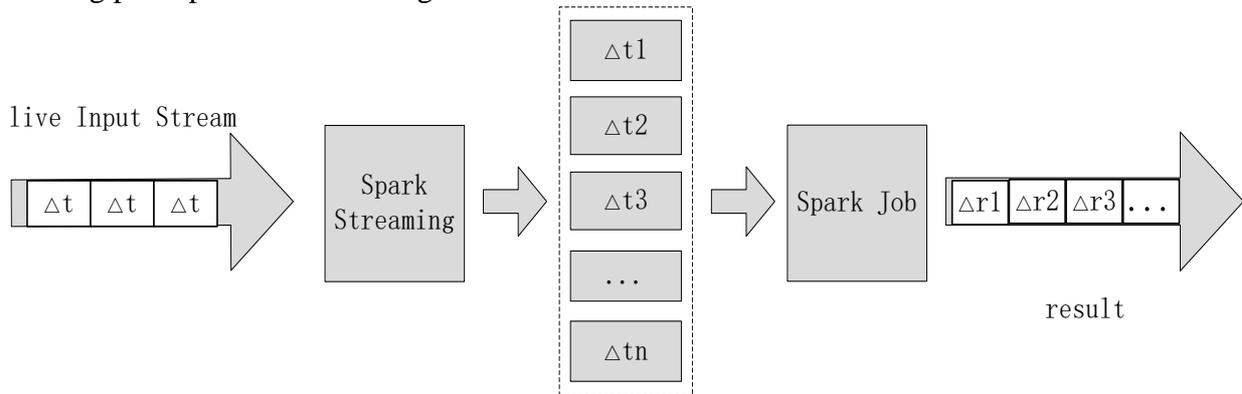


Fig.1. The working principle of Spark Streaming

The fast fault diagnosis framework for electric metering device based on Spark Streaming

Stream processing technology can provide effective support for the real-time processing of the information of the smart grid, and has broad prospects for research. According to the smart grid's characteristics, i.e., real time, volatile and disorder, a fast fault diagnosis framework based on Spark Streaming is proposed, which is shown in Figure 2. The framework mainly consists of five parts: data acquisition, data access, stream processing, data storage and data presentation.

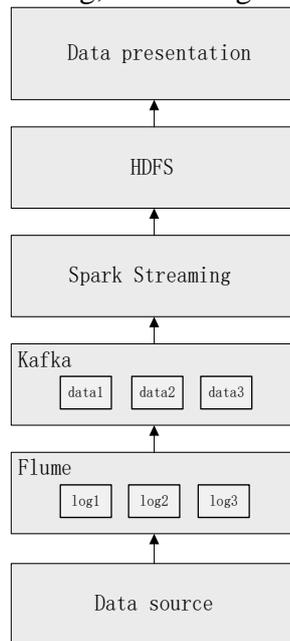


Fig.2. The fast fault diagnosis framework for electric metering device based on Spark Streaming

This paper uses Flume as the real-time data acquisition system to collect the terminal data from the electricity measuring device. Flume is a mass log collection, aggregation and transmission system, which is distributed, reliable, and highly available. Flume supports a variety of customized data sender in the log system. The data can be processed simply and written to the receiver. Kafka is a high throughput distributed message distribution and subscription system. It can classify data according to the topic, and then send data to Spark Streaming. Flume is more focused on data transmission, while Kafka is a typical message middleware. Data is processed in Spark Streaming and then Spark Streaming can save the results to HDFS for further data presentation.

Experiments and results analysis

In this paper, the data collected by the power consumption information collection system is used as the input data, and simulated as a real-time data stream on the cluster. There are many kinds of diagnosis methods: neural network, support vector machine, expert system, Naive Bayesian classifier and so on. In this paper, BP neural net work algorithm is used for fault diagnosis for metering devices.

In this paper, six computers are used to build the Spark Streaming cluster. One is used as the master node, and the other five computers are used as worker nodes. The nodes in the cluster have the same configuration: 2G memory, 60G hard disk, 64 bits Centos operating system. The experiment is also carried out in stand-alone environment. With the following machine configuration: 2G memory, 60G hard disk, 64 bits windows7 operating system.

The collected data is simulated as a data stream. Experiments are carried out in stand-alone environment and cluster environment. The comparison of test results is shown in Figure 3.

As shown in Figure 3, the processing performance of the stand-alone mode is better than that of the cluster mode for small data, which is due to communication overhead, network delay and working deployment time in the cluster. With the increase of the amount of data, the effect of stream processing cluster is better than the stand-alone mode, which can achieve a higher real-time performance.

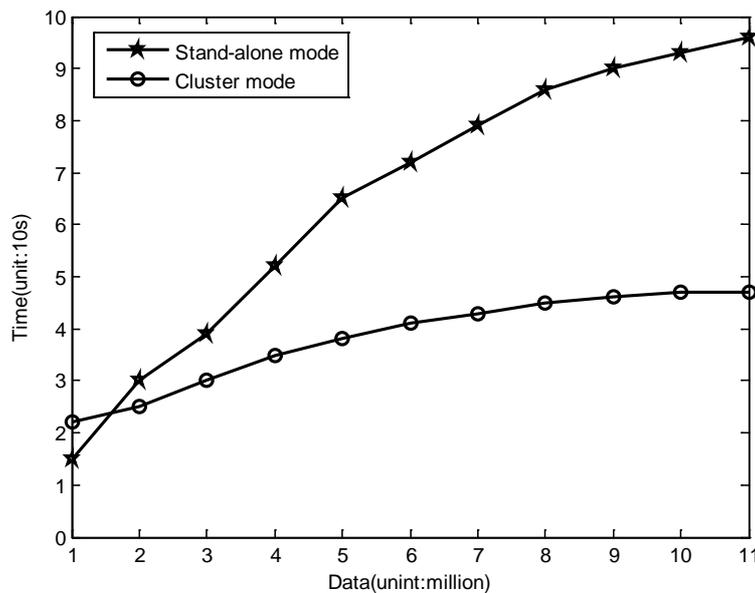


Fig.3. Spark Streaming cluster and stand-alone mode run time comparison

Summary

This paper analyzes the importance of real-time processing of the data stream generated by the metering devices, and proposes a fast fault diagnosis framework for electric energy metering device based on Spark Streaming. The data collected by power consumption information collection system is used as the experimental input data, which is simulated as a real-time data stream. BP neural network algorithm is used to process the data stream to diagnose the faults. The experimental results show that the performance of Spark Streaming cluster is better than that of traditional stand-alone mode. And it has wide application prospect in the field of fault diagnosis of electric energy metering device of smart grid.

References

- [1] Chang Qing, Yan Xiaowen, Tao Xiaoyan, Fu Feng. Research on the operation state analysis system of intelligent electric meter based on big data technology[J]. Automation and Instrumentation. 2015, 12: 4-6+9.
- [2] Wang Dewen, Yang Liping. Big data stream processing method and state monitoring anomaly detection for smart grid[J]. Automation of Electric Power Systems. 2016, 14: 122-128.
- [3] ZHANG Shaomin, MAO Dong, WANG Baoyi. Application of big data processing technology in fault diagnosis and early warning of gearbox of wind turbine generator[J]. Automation of Electric Power Systems. 2016, 14: 129-134.
- [4] Lou Qian. Research and development of electricity meters online monitoring and diagnosing system[D]. North China Electric Power University. 2015.
- [5] Cao Min, Liang Shibin, Li Yi, Wang Xin. Study on on-line monitoring system of electric energy metering device in yunnan power grid[J]. Electrical Measurement and Instrumentation. 2007, 03: 40-42+54.
- [6] Fan Jie, Chen Xiao, Zhou Yu. Research on abnormal intelligent analysis method of electric energy metering device based on electricity information collection system[J]. Electrical Measurement and Instrumentation. 2013, 11: 4-9.
- [7] Li Youkun. Analysis and improvement of BP neural network[D]. Anhui University of Science and Technology, 2012.