Advances in Intelligent Systems Research (AISR), volume 145
2017 International Conference on Electronic Industry and Automation (EIA 2017)

A Stream Computing Method for Railway Distribution Network Monitoring Information Based on Storm

Zhijian QU
School of Electrical and Automation Engineering
East China Jiaotong University
Nanchang, Jiangxi Province, China
e-mail: 08117324@bjtu.edu.cn

Qunfeng WANG
School of Electrical and Automation Engineering
East China Jiaotong University
Nanchang, Jiangxi Province, China
e-mail: 495834944@qq.com

Xiang PENG
School of Electrical and Automation Engineering
East China Jiaotong University
Nanchang, Jiangxi Province, China

Ruilin ZHOU
School of Electrical and Automation Engineering
East China Jiaotong University
Nanchang, Jiangxi Province, China

***Abstract*—Aiming at the problem of rapid processing of real - time data in intelligent dispatching, a new method for calculating the information flow of distribution network monitoring is proposed. Access the distribution network monitoring data by publishing subscriptions, combining the topology parallel model of flow calculation, comprehensive use of multi - theme partition message caching technology, to achieve low-latency and high-throughput processing of distribution network monitoring information[1-3]. The method can obtain hundreds of millisecond monitoring data processing delay, and it has practical value in research to improve low-latency distribution network dispatch system.**

Keywords-intelligent dispatching; stream computing; publish subscribe; Storm

I. INTRODUCTION

With the intelligent distribution network in a large number of intelligent electronic equipment applications and panoramic data acquisition needs, power grid operation required for remote, telemetry and alarm information increased significantly. According to a regional grid automatic dispatch system reported in the literature, when the grid anomalies or failures, in a short time will produce hundreds or even thousands of alarm information into the control center, these alarm messages are presented directly to the dispatcher in a rapidly changing form, the dispatcher should select important warning information to deal with the mass of alarm messages, it is a difficult and time-consuming job[4-6]. So if not timely processing of these massive monitoring information, it may delay the processing time of the fault, and even lead to security incidents.

Big data processing technology can be divided into two kinds of batch calculation and flow calculation, the first kind of batch calculation is stored first and then calculated, it is suitable for large-scale data distributed storage and off-line processing. Take the Hadoop batch calculation framework as the representative, since Hadoop first distributes the data distribution to the data node cluster, then batch processing the node data in the form of data slicing, this process is not conducive to a large number of real-time monitoring of data processing. The second category is the new stream computing model represented by Storm, stream computing uses distributed memory computing to avoid memory capacity limitations, using Storm distributed real-time flow computing model can provide new technical means to solve the fast response of mass monitoring data in distribution network. This paper considers the distributed message queue to be integrated into the data extraction link of the stream computing cluster, in order to improve the throughput and response speed of mass monitoring data in distribution network.

By building Storm stream computing cluster, establish a distributed message cache queue for real-time monitoring of data, using the topic partitioning mechanism and the stream computing parallel topology programming model, to achieve rapid processing of mass data. Load actual engineering data as a data source, stream analysis of distribution network monitoring information, verified that the proposed method has low delay characteristics in mass monitoring data processing.

II. DISTRIBUTION NETWORK MONITORING DATA CACHE MODEL

A. Message Queuing System Partition Mechanism

The message queue system uses partition, segmentation and sparse index to ensure the efficiency of message transmission[7-8]. The partitioning mechanism of a distributed message queue is stored as a file in a file system. Each message in the partition contains three attributes: offset, message size, and metadata. In a partition of a distributed message queue, a large message file is equally allocated to several files of equal size, that is, a large file is cut into several equal sized file segments. The file segment consists of two parts: the file segment data file and the file segment index file. The two parts are the one-to-one relationship.

Copyright © 2017, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

B. Real-time Monitoring of Data Access Mode

In a message queue system usually includes multiple cache agents, in order to achieve load balancing, the message theme is divided into multiple partitions, each cache agent stores multiple partitions, multiple producers and consumers can simultaneously produce and get the message. The message queuing system manages messages with topics, each topic contains multiple station partition, each station corresponds to a logical partition log. Each partition consists of a number of segments, each of which stores a certain amount of message. After the distribution network monitoring data is sent to the corresponding topic in the message queue, the message is evenly distributed to multiple partitions. After receiving the monitoring message from the cache node, the message is added to the last segment of the corresponding station partition. When the number of messages on a segment reaches the preset value or the message publication time exceeds the threshold, the message in the segment is written to disk. Message queue access architecture is shown in Fig. 1.

III. Distribution network monitoring data stream processing mode

A. Storm Parallel Topology Programming Model

Storm is an open source distributed flow computing framework, using flow computation topology and parallel programming model to design real-time data processing process as a directed acyclic topography[9]. The distribution network monitoring real-time data stream is abstracted as uninterrupted continuous tuples, each data stream has a data source, that is the source of the original tuple. The user-created spout assembly can continuously read the monitoring data from the message queue to assemble as a tuple to form a tuple data stream. Using a bolt component to define the computing operations of tuples in a data stream, it is possible to perform the operation of decomposing aggregated data, continuous calculation and so on, and sending the new tuple to the next data logical processing bolt in the topology. When the data spouts and bolts processing components in the flow calculation cluster are set, the computational tasks are submitted in topology to the flow calculation cluster. Storm parallel topology model is shown in Fig. 2.

B. Distributed Publishing Subscription Architecture

Kafka is a distributed message queue that uses message queues to persist data to avoid the risk of data loss. The telemetry information and remote information collected by the dispatching center are first distributed as a message producer to the cache node. In a message cache node contains a leader node and multiple followers node, in the leader node and follower node, respectively, store the same theme of the different partitions, between the leader node and the follower node can be synchronized a copy of the monitoring data. When the dispatch center sends telemetry information and telemetry information to the distributed message queue, the spout and bolt components in the flow calculation cluster are used as message consumers to extract the monitoring data from the subject of the message queue for flow calculation. In order to run the entire flow computation cluster in an orderly manner, the message producers, message cache nodes, and message consumers in the cluster register the status values into the Zookeeper cluster so that the Zookeeper cluster coordinates the global resources of the message queue. Stream computing cluster architecture is shown in Fig. 3.

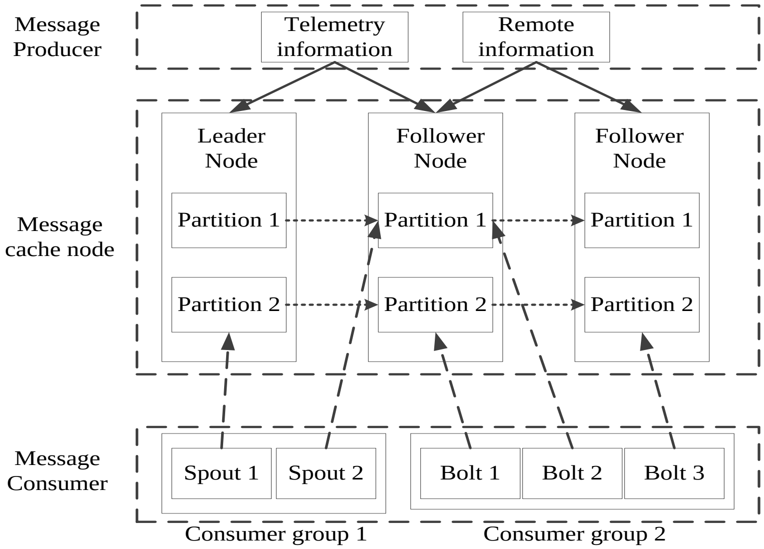

Figure 1. Message queue access architecture

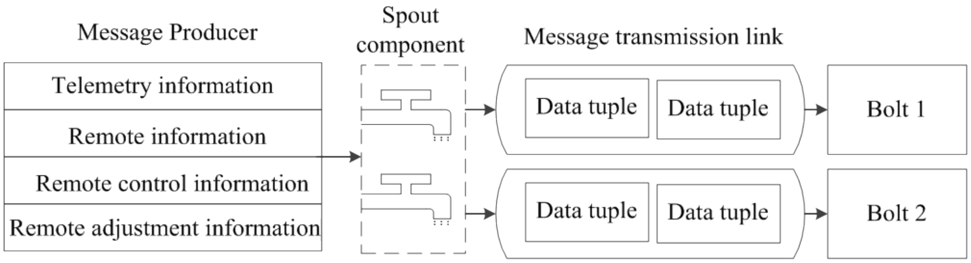

Figure 2. Storm parallel topology model

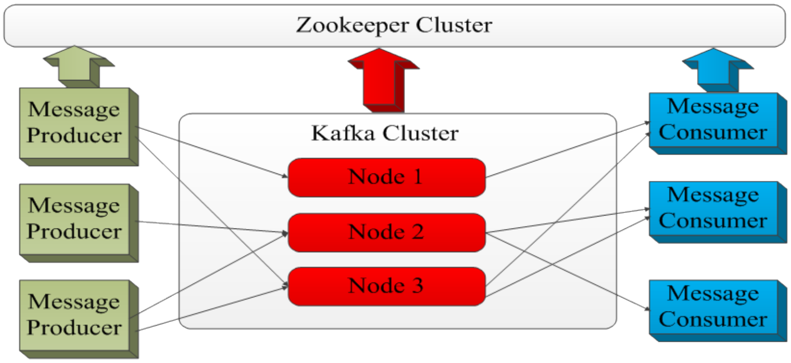

Figure 3. Stream computing cluster architecture

ACKNOWLEDGEMENT

Project Supported by the National Natural Science Foundation of China (51567008); Foundation Plan for Distinguished Young Scholars in Jiangxi Province (20162BCB23045); Natural Science Foundation of Jiangxi Province (20161BAB206156, 20171BAB206044); Science and Technology Research Project of Jiangxi Provincial Education Department (GJJ160471).

REFERENCES

[1] Song Yaqi, Zhou Guoliang, Zhu Yongli. Present Status and Challenges of Big Data Processing in Smart Grid[J]. Power System Technology, 2013, (04): 927-935(in Chinese).

[2] Liu Keyan, Sheng Wanxing, Zhang Dongxia, Jia Dongli, Hu Lijuan, He Kaiyuan. Big Data Application Requirements and Scenario Analysis in Smart Distribution Network[J]. Proceedings of the CSEE, 2015, (02): 287-293(in Chinese).

[3] Zhang Dongxia, Miao Xin, Liu Liping, Zhang Yan, Liu Keyan. Research on Development Strategy for Smart Grid Big Data[J]. Proceedings of the CSEE, 2015, (01): 2-12(in Chinese).

[4] Zhao Dongmei, Zhang Xu, Zhang Dongying, Liu Yanhua, Wei Juan, Zhang Hong. Basic Analysis Modes of Fault Diagnosis for Regional Power Grid[J]. Power System Technology, 2012, (11): 184-191(in Chinese).

[5] Wu Kaifeng, Liu Wantao, Li Yanhu, Su Yipeng, Xiao Zheng, Pei Xubing, Hu Songlin. Cloud-Computing Based Power Big Data Analysis Technology and Its Application[J]. Electric Power, 2015, (02): 111-116+127(in Chinese).

[6] Wang Dewen, Yang Liping. Stream Processing Method and Condition Monitoring Anomaly Detection for Big Data in Smart Grid[J]. Automation of Electric Power Systems, 2016, (14): 122-128(in Chinese).

[7] Jia Chaolong,Wang Hanning,Wei Lili. Research on Visualization of Multi-Dimensional Real-Time Traffic Data Stream Based on Cloud Computing[J]. Procedia Engineering,2016,137(1).

[8] Pan Guowei, Song Wei, Wang Xiangnan, Cao Houji. Research on Application of Message Oriented Middleware Based on Publish/Subscribe Model in SCADA System[J]. Power System Technology, 2008, (18): 77-81(in Chinese).

[9] Wang Zhen, Chen Liang. Quasi-real-time Data Access Method of Power Network Equipment Based on Kafka Message Queue[J]. Shandong Electric Power, 2015, (06): 41-43(in Chinese).