

Design and Key Calculation of A Big Data Platform for Transmission Channels Monitoring

Rui Liu

*Transmission Line Operation and Maintenance Branch of Guizhou Power Grid Co., Ltd
Guiyang 550005, China
E-mail: 73495435@qq.com*

Tao Guo^a, Ping Shen^b

*Transmission Line Operation and Maintenance Branch of Guizhou Power Grid Co., Ltd
Guiyang 550005, China
E-mail:^a79169957@qq.com;^b1054996914@qq.com*

Yue Chen^{1c}, Lei Zhang^{2d}

*Wuhan NARI Limited Liability Company, State Grid Electric Power Research Institute,
Wuhan 430074, China
E-mail:^c597265154@qq.com;^d240466862@qq.com*

You Hu^e, Fei Luo^{3f}

*Computer school, Wuhan university,
Wuhan 430074, China
E-mail:^e2014301500106@whu.edu.cn;^fluofei_whu@126.com*

During the period of China's 13th five-year plan, the investment in power grid continues to increase. By 2020, a strong smart grid will have been built up and the goal of building a world-class power grid will have been achieved. Monitoring the state of transmission channels could guarantee power grid's stability and reliability. In order to better manage power grid, it is necessary to improve the ability of processing and analysing data. In this paper, we proposed a big data platform to monitor the state of transmission channels. The mapping operation between lighting position and the pole's position needs great computation as lightning is a typical monitored variable. In the real data experiment, MapReduce has a significant improvement in speed compared with the traditional method.

*CENet2017
22-23 July 2017
Shanghai, China*

¹Corresponding Author

²Speaker

³This study is supported by the National Science Foundation of China (61402340, 61572368).

1.Introduction

In the power industry, the big data has been promoted to a strategic height. In 2013, the China Electric Engineers Electric Power Information Professional Committee issued the "China Power Big Data Development White Paper", which pointed out the inevitability that the big data should be developed in the power industry. In 2015, "the big data application guidance in State Grid Corporation" was issued. It proposed the top-level planning for the big data application, including three major areas of 35 typical application scenarios. In recent years, a number of successful big data pilot cases have emerged in the power industry. In one paper[1], the big data technology was used to excavate the huge data of the power grid with efficiency and rapidity, and realized the power grid online safety evaluation. Combined with the cloud computing, work [2] proposed a big data platform on power consumer side, which integrated data from the smart ammeter, SCADA system and a variety of sensors, and utilized both MapReduce and memory computing Spark. Work [3] designed and developed a big data storage system based on the Hadoop, including Hadoop cluster, storing client and querying client. To meet the demand of real-time analysis of the power system, work [4] discussed the solution on the core problems such as power system security, real-time state monitoring and dynamic energy balance scheduling under the circumstance of big data.

Although the big data has been widely used in power industry, such as smart power grid, security, data storage, and power grid operation, there are few researches aiming to explore the transmission channels monitoring. China has vast territory, across which the meteorology and terrain conditions of transmission lines change greatly, so many components' data of transmission lines need to be monitored, including the wires, poles, fittings and insulators.

The difficulties to effectively monitor and manage the transmission channels come from many aspects. For example, in aspect of hardware, due to the variety of devices on the transmission channels, it lacks the standard and uniformed sensor and interface to simultaneously monitor all of them [5]. In terms of software, there are a number of systems that have been proposed to monitor specific factors affecting the safety of transmission lines, such as lightning monitoring [6], windborne monitoring [7], icing monitoring [8], contamination monitoring [9], fire monitoring [10] and so on, but the systems are independent to each other and shortof effective information sharing and sustainable development.

These difficulties will remain unchanged for a long time, and improving the efficiency of the power grid's operation and maintenance will be a keeping-on goal. Huge investment in upgrading the hardware in a short-time is unrealistic, but from the software side, it's possible to make overall plans and take all factors into consideration at the start, and to be flexible enough to adjust scheme during the implementation stage. Therefore, it is an important breakthrough to establish a comprehensive data analysis system for the power grid operating and maintenance by taking advantage of the big data [11]. Based on the big data technology, this work proposes a big data framework for integrating, managing and utilizing the transmission channels' monitoring data, and offers a MapReduce method for key calculation of mapping typical monitored quantity to the corresponding power equipment.

2.Big Data Framework for the Transmission Channels Monitoring

Hadoop, Spark, and Storm are the three major solutions which are widely used in the big data technology. Hadoop consists of the distributed computing technique MapReduce, distributed file system HDFS and data storage system HBase. Spark possesses some structure improvements over the Hadoop. Comparing to Hadoop’s using hard disk to store data, Spark stores data in RAM memory, which provides Spark with faster computing speed. But Spark can’t process data that requires permanent storage. Storm can be viewed as a real-time computing system. Thus Storm is suitable for processing stream data and can be used to deal with a steady stream of incoming messages. From the scenarios that three technologies are suitable for, Hadoop is often used for complex off-line processing. Spark is used for fast off-line processing, while Storm is used for real-time online processing.

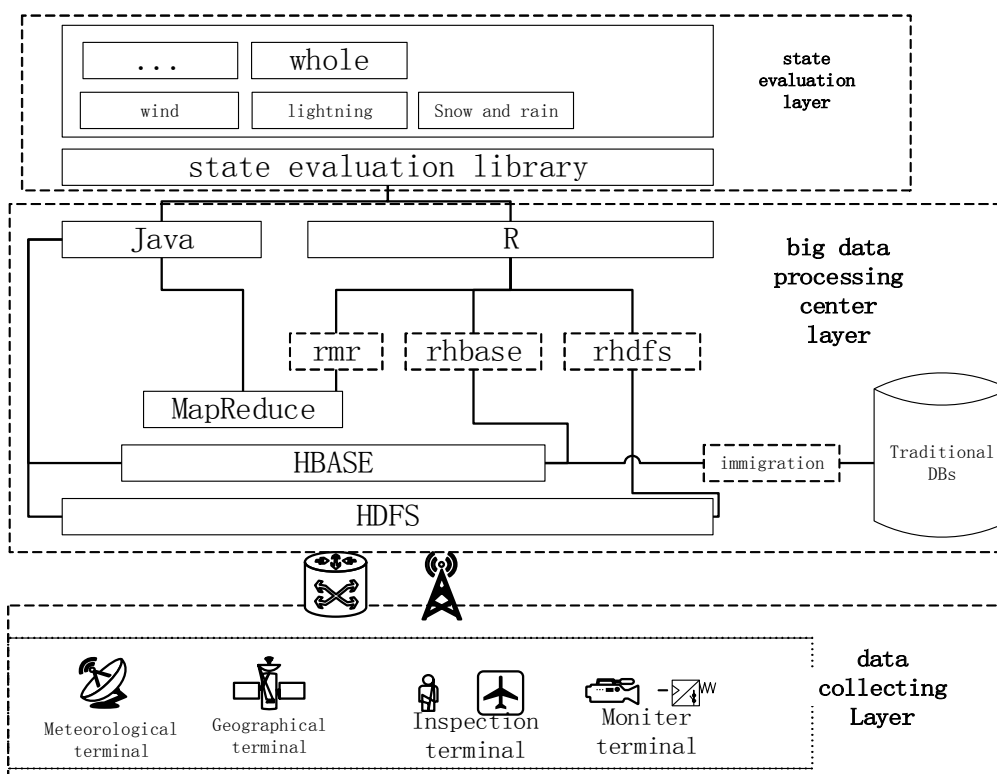


Figure 1: Big Data Platform Framework for Monitoring the Transmission Channels

Hadoop actually becomes the standard for distributed computing systems. Especially it is suitable for long-term data preservation and analysis. Therefore, Hadoop is used here to propose a big data platform for transmission channels monitoring. Figure 1 is a schematic diagram of the framework which consists of three layers: the terminal data collecting layer, the big data processing center layer, and the transmission channels’ state evaluation layer.

The data collecting layer collects various basic information to ensure the final transmission channels’ state evaluation is truly reliable. Meteorological factors are major external threatening forces for the transmission channels. Through weather radar, thermometer, anemometer, hygrometer, pressure meter and other devices, meteorological data are collected. Through the GPS and satellite data, geographic data are provided with the information about geographical environment where transmission channels locate. Depending on unmanned aerial vehicles and other robots, the inspection data timely reflects the state of specific part of the transmission channels. Through cameras and voltage sensors, leakage current sensors, vibration sensors and other sensors, the lasting monitoring provides continuous information of important part of

transmission channels. All terminals' data are sent to the big data processing center through wired or wireless network.

The big data processing center uses HDFS and HBASE to provide high-throughput data visiting ability. There are two kinds of data need to be stored. One is the dynamic data from the terminals. The other is the transmission channels' components information, which is usually stored in traditional databases like SQL Server or Oracle. Hadoop has the module to immigrate such data. MapReduce in the big data center play a key role in providing a parallel computing model. It can quickly and efficiently perform computation in a way of splitting, mapping and merging. In order to write MapReduce procedures, big data center provides Java and R programming language interfaces. Java is the standard programming language in Hadoop, and R is the statistical language, which is good at mathematical statistics, machine learning. Therefore, the center can conveniently conduct transaction processing, mathematical statistics and machine learning programming. The R tightly connects with MapReduce, HDFS and HBASE through rmr, rhdfs and rhbase components.

Based on association rules and principal component analysis [12], there are many methods for the evaluation of transmission channels' state, such as, the delphi method and hierarchical model [13], simple parameter set [14] and so on. According to the past experience, the single evaluation method can hardly fully meet the demand because the condition is always changing. So, instead of limited to a certain fixed evaluation method, the transmission channels state evaluation library is designed in the state evaluation layer, and any new evaluation algorithm could be added into library and freely modify or remove those existing ones according to the needs of the task to make the evaluation more scalable and comparable.

3. MapReduce Solution for the Key Calculation

No matter what kind of evaluation method is used, the evaluation of the transmission channels' state must depend on certain critical monitoring variable such as the number of lightning occurrence, wind strength and the number of defects and so on. When counting these variables and analyzing how they distribute along the channel, it is necessary to map the monitored variable to its nearest equipment. This kind of mapping operation needs huge computation and is the fundamental for subsequent evaluation of these transmission channels.

```

Process 1:
class Mapper
  method Initialize
    thunderLength ← numberOfThunders // initialize the lightning's number
    poleLength ← numberOfPoles // initialize the pole's number
    CountT ← 0
    CountP ← 0
  method map(docid a, doc d)
    IF file match "thunder" // when data comes from the lightning
      CountT ← CountT+1
      for I in 1:poleLength
        (key,value) ← {{CountT,i},{long,lat}} //generate the key-value pair
        Emit(key,value) // emit to Reduce
    ELSE IF file match "pole" //when data comes from the pole
      CountP ← CountP+1
      for I in 1:thunderLength
        (key,value) ← {{i,CountP},{long,lat}}
        Emit(key,value)
class Reducer
  method Reduce(term t, values[v1,v2])
    distance ← computeDistanceOf [v1,v2] // compute the distance between lightning and pole
key ← extract thunderId From t //compute key
poleId ← extract poleId From t //extract the pole's index
value ← {poleId,distance} //generate value
    Emit(key,value) //emit, prepare for next MP process

Process 2:
class Mapper
  method map(docid a, doc d)
    (key,value) ← extract from d //preprocessing
    Emit(key,value)
class Reducer
  method Reduce(term t, values[p1,p2,p3,...])
    dn ← extract distance from [p1,p2,p3...] // extract the shortest distance
IF dn ~=NULL
poleId ← poleIdOf (dn) //extract pole id from the lightning and pole pair meeting
requirement
Emit(poleId,1)

process 3: //calculate the number
class Mapper // preprocessing
  method map(docid a, doc d)
    (key,value) ← extract from d
    Emit(key,value)
class Reducer
  method Reduce(term t, counts[c1,c2,c3,...])
    sum ← sum[c1,c2,c3,...] //count
    Emit(t,sum) //final result

```

Table 1: MapReduce Pseudo Code

Therefore, lightning, the most typical and important monitored variable for the transmission channel, is taken as an example to illustrate this process. A MapReduce method is proposed to realize the quick mapping (shown in Table 1). The task is to use the locations of lightning occurring and pole's position to count the lightning number of each pole with the shortest distance principle. After mapping, the distribution of lightning along the line during a period could be revealed, which indicates high-risk poles of transmission channels and needs further protection.

Based on the data of one year, the lightning data set can be cut into many small subsets by MapReduce, each of which runs mapping operation in parallel to each computing node to count statistics. After the statistics are completed, the results are merged and summarized. In particular, the process is essentially one joining operation for two sets.

4. Experiment

In the experiment, Hadoop-2.6.0 is used to build the experimental environment under ubuntu 14.04LTS. Hadoop has three modes of figuration: Standalone mode (native mode), daemons, and JVM. Testing and debugging in stand-alone mode is easy and commonly used for the development stage; pseudo-distributed mode runs daemon locally and simulates a small-scale cluster; full distributed mode runs daemon on a cluster. This experiment takes pseudo-distributed mode.

Three transmission channels are used in Xinjiang, named ACA, BTB2 and CKA. They have respectively 413, 712 and 404 poles. The time of lightning monitoring data starts from 2013-01-01 00:00:00 to 2014-01-01 00:00:00, total 463675 times lightning. In order to exclude lightning which are too far, the distance between the lighting and corresponding pole is required to be less than 3km.

Finally MapReduce method shown in Table 1 gives the lightning distribution along three lines shown in Figure 2. The lightning for ACA is mainly concentrated on the 80 to 150 and 350 to 413 poles; the lightning for BTB2 mainly concentrates on the 200 to 270 poles; the lightning for CKA lies between 300 to 330 poles. In order to verify the effectiveness of the method, the same distribution is shown in Figure 2 by using the same distance matching method but realizing with the traditional serial Java programming on the same machine . It proves the correctness of the MapReduce. Because MapReduce splits raw data for parallel processing, the time spent is only 1/3 of the traditional Java method.

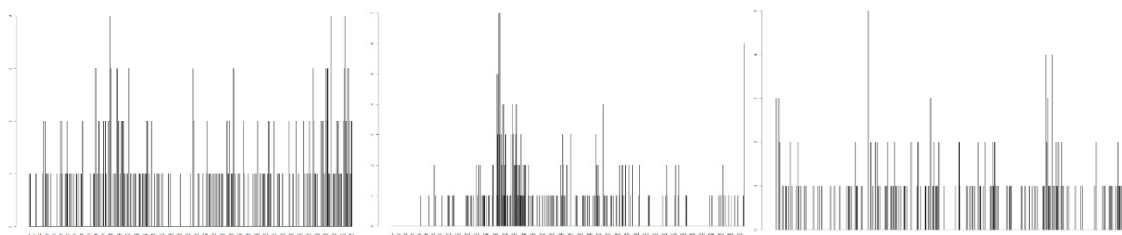


Figure 2: Key Mapping Calculation Results by MapReduce

5. Conclusion

The power industry is an important application direction of the big data technology. With the gradual maturation of power grid construction and the requirement of intelligent and efficient managing, the demand for power grid maintenance is increasing. This paper demonstrates a framework of the big data platform for transmission channels monitoring. At the same time, it discusses the key calculation and realizes its implementation with MapReduce.

References

- [1]Feng Changyou, Jin Yiding, Shi Dongyu, Zhen Xiaoyan, Wang Zhen, Qiu Wei.*Remote Unified Computing Techniques for Inter-grid Online Security and Stability Analysis and its Application Based on D5000 Smart Grid Dispatching and Controlling System*[J]. Shanxi Electrical Power.43(2): 39-44(2015) (In Chinese)
- [2]Wang Dewen, Sun Zhiwei.*Big Data Analysis and Parallel Load Forecasting of Electric Power User Side*[J]. Proceedings of the CSEE. (3):527-537(2015) (In Chinese)
- [3]Liu Shuren, Song Yaqi, Zhu Yongli, Wang Dewen. *Research on Data Storage for Smart Grid Condition Monitoring Using Hadoop*[J]. Computer Science. 40(1):81-84(2013) (In Chinese)

- [4]Zhou Guoliang, Lv Linjie, Wang Guilan. *Key technology of power big data for global real-time analysis*[J]. Telecommunications Science. 32(4):159-168 (2016) (In Chinese)
- [5]Cao Yijun, Wang Yining, Qian Hao, Ma Guilin, Wei Minxiang, Dai Jiayuan. *A general solution for on-line inspection of transmission lines*[C]. China electric power technology collaboration (EPTC)2014 the third conference.454-458(2014) (In Chinese)
- [6]Zhang Qiming, Zhou Ziqiang, Gu Shanqiang, Feng Wanxing, Wang Haitao. *Cloud Computing Application Technology of Massive Lightning Monitoring Data*[J]. Automation of Electric Power Systems. 36(24):58-63(2012) (In Chinese)
- [7]Hu Zhijian, Li Hongjiang, Wen Xishan, Zhang Chengxue. *Online monitoring of transmission conductor wave and wind gallop based on differential GPS*[J]. Electric Power Automation Equipment. 32(3):120-124(2012) (In Chinese)
- [8]Wang Xiaoqing, Dai Dong, Hao Yanpeng, Li liying, Fu Chuang, Rao hong. *Statistics and Analysis of Transmission Lines Icing Data Based on Online Monitoring System*[J]. High Voltage Engineering.38(11):3000-3007(2012) (In Chinese)
- [9]Jiao Shangbin, Liu Ding, Zhen Gang, Li qi. *On-Line Insulator Contamination Monitoring System For Transmission Lines Based On Telemetry*[J]. Automation of Electric Power Systems. 28(15):71-75,94(2004) (In Chinese)
- [10]Ye Liping, Chen Xiyang, He Zhilan, Xie Congzhen, Huang Jianhua, Xia Yunfeng. *Present situation of forest fire early warning technology used for transmission line*[J]. Power System Protection and Control. (6):145-153(2014) (In Chinese)
- [11]Hu Yi, Liu Kai, Wu Tian, Liu Yan, Su Zhimin. *Analysis of Influential Factors on Operation Safety of Transmission Line and Countermeasures*[J]. High Voltage Engineering.40 (11):3491-3499(2014) (In Chinese)
- [12]Yan Yingjie, Sheng Gehao, Chen Yufeng, Guo Zhihong, Du Xiuming, Wang Qi. *Establishment of Key Parameter System for Condition Evaluation of Transmission Line Based on Association Rules and Principal Component Analysis*[J].High Voltage Engineering. 41(7):2308-2314(2015) (In Chinese)
- [13]Wang Chengjiang, Li Hongyan, Fan Kaiming, Zhang Boya, Li Guang. *State Evaluation of Power Transmission Line with Delphi Method Based on Hierarchical Model*[j]. Journal of China Three Gorges University (Natural Sciences).34(5):52-57(2012) (In Chinese)
- [14]Song Yunhai, Chen Yue, Li Jinwei, He yu. *Research on Transmission Line State Evaluation Method Based on Simple Parameter Set*[C]. 2015 proceeding of Chinese society for electrical engineering. 1-7. (2015) (In Chinese)