A CEP based ETL method of active distribution network operation monitoring and controlling signal data

Minghao AI
CEPRI-China
aiminghao@epri.sgcc.com.cn

Naishi CHEN
CEPRI-China
chennaishi@epri.sgcc.com.cn

Xianjun GE
CEPRI-China
gexianjun@epri.sgcc.com.cn

Zhihong LI
CEPRI-China
lizhihong@epri.sgcc.com.cn

Tianjiao PU
CEPRI-China
putianjiao@epri.sgcc.com.cn

Ye LI
CEPRI-China
liye@epri.sgcc.com.cn

Zhengping CHEN
State Grid Fujian Electric Power Company-China
czp820118@163.com

Pi LIN
State Grid Fujian Electric Power Company-China
13606081702@139.com

Wei WU
State Grid Fujian Electric Power Company-China
543619893@qq.com

ABSTRACT

With the development of Active Distribution Network, the scale of power system becomes larger and larger, and the number of electrical equipment in distribution network increases sharply and becomes further precise, electrical equipment's operation monitoring and controlling signal data has the characters of massive, diversity and complication, shows a trend of Big Data. Massive and random operation monitoring and controlling signal data causes various applications in active distribution network unable to extract useful information quickly and efficiently so as difficult to form decision support. Complex event processing (CEP) is an intelligent data processing technology rise in the era of Big Data, which can implement rapid analysis and processing to continuous data based on rule engine. The article uses CEP engine as the operation monitoring and controlling signal processing core, and uses ETL (Extract-Transform-Load) framework to integrate, clean and load the distributed, disordered and standard not unified signal data in active distribution network into the data warehouse. The problem of data format not unified and independent storage during data extraction can be solved by using the adapter mode and daemon process way. Based on CEP engine, it determines the core processing architecture of operation monitoring and controlling signal big data. In the architecture, signal cleaning rule library and algorithm library use the pluggable mode which makes them easy to maintain and expand. Rules library can be determined by using nested query, combined operation and pattern matching, and algorithm library can be packaged of memory partitioning and multithread processing, word-frequency statistics, keyword recognition and elimination, and other algorithms. It uses buffer queue to cache processing result and format the output as needed. The CEP engine based Big Data ETL solution implements the fast, accurate and effective standardization processing of operation monitoring and controlling signal and provides accurate data preparation for fast simulation, fault analysis, state estimation and other important application in active distribution network.

Keywords: *CEP Engine, Operation Monitoring and Controlling, Big Data, ETL, Active Distribution Network, Standardization*

INTRODUCTION

Because of power grids around of China have used different manufacturers, different types of electrical equipment, the format and naming of operation monitoring signal produced by various devices are also various kinds and huge amount. This makes each power system applications based on the signal and information systems integration in the process of building the World Energy Internet encountered great challenges, including grid state estimation, scenario analysis, signal driving simulation and other actual demands. In addition, there is no unified standard of substation secondary equipment and systems operation monitoring signal formats and naming of substation secondary equipment and system operation monitoring signal format and naming, so that clutter signals can't be unified named as a basis for data store[1]. Through research, at home and abroad, there isn't an effective treatment method for electrical equipment operation monitoring signal standardization, let alone actual cases of signal standardization processing.

ETL (Extract-Transform-Load)[2] is the process of loading the business system data into the data warehouse after extracting, cleaning and transformation, whose purpose is to integrate the spread, messy, non-unified standard data together, so that provide data to prepare for the subsequent application.

CEP (Complex Event Processing)[3] is an intelligent data processing technique rising in the Big Data Era, which regards the data as a data stream, complex processes various events continuously generated in the business process based on rule engine, so that implements the rapid analysis and handling of continuous data. The

CEP engine can solve the problem of transactional data processing efficiency and traditional MQ[4] way difficult to deal with the issue of complex events. So it is particularly suited to the requirements of adopting rules library and intelligent learning algorithm to process big data. For the secondary equipment[5] operation monitoring signal data standardization work, we can based on the prepared rules libraries, using pluggable type intelligent algorithms library, integrated to the CEP engine for processing, so that to achieve significant effect.

EXTRACTION OF OPERATION MONITORING SIGNAL BIG DATA

Distribution network operation monitoring signal storage

In the current distribution power automation system, equipment operation monitoring signal is stored mainly in four forms: real-time database, historical database, common exchange files and NoSQL database.

Signal data stored in the real-time database and historical database adopts traditional relational database management mode, the difference is physically stored in memory and hard disc respectively, which caused different access efficiency and durability and applies to different real-time and data volume scenarios. Table 1 shows the main content of operation monitoring signal table stored in this way.

Tab.1 Relational database signal storage format

FieldName	FieldDescription	FieldType
id	Unique identify	long[64]
name	Signal Chinese name	char[128]
st_id	Substation identify	long[64]
vl_id	voltage grade identify	long[64]
pnt_type	Signal type	char[1]
oo_dev	Belong device	long[64]
pri_flags	Protection priority	char[1]
brk_num	Breaker number	int
…	…	…

In order to facilitate the migration, transformation and data access universality of operation monitoring signal, some applications using the common exchange file format to access the distribution network operation monitoring signal data. XML format is an international common exchange data format with the advantages of good compatibility, flexible and extensible. Based on the XML standard, according to its strong extensibility, for power system characteristics, the CIM/E standard is put forward by SGCC (State Grid Corporation of China), which dedicated to online exchange of power grid model data among each application system in dispatch control center. Operation monitoring signal data CIME file storage format is as follows:

<relaysig>
@id name vl_id pnt_type pri_flags …
#122160141620511071 overcurrent 2nd plate 0001 1 3
</relaysig>

With the expansion of the distribution network scale and increasingly promoted of the complexity, operation monitoring signal data quantity is rising exponentially, showed a trend of big data. The traditional relational databases and file storage way have been unable meet the demand of reading efficiency and so on. Therefore, operation monitoring signal storage will develop in the direction of large data storage. The NoSQL storage such as HBase column-store database can guarantee the mass signal data query efficiency, history library, data scalability storage requirements, etc.

Since the data source of current power distribution automation system around is relatively single, the data are stored separately and do not affect each other. For example, operation monitoring signal data of 1 province and 9 regions in Fujian province is distributed storage in 10 data server such as DTS-FJ(Fujian) 、 DTS-QZ(Quanzhou)、DTS-LY(Longyan) and so on, without interference of data operation to each other. This ensured the independence and security of the data, but caused difficulties to global data extraction and analysis.

Operation monitoring signal on-demand extraction

In order to achieve directional, custom signal analysis service, we need to filter and extract operation monitoring signal data according to each field, regional, substations, voltage levels and other conditions. Through the analysis of the operation monitoring signal data storage status in distribution network, we will face two kinds of problems to realize the on-demand extraction of signal data:

(1) Multi-format data problem. Current signal data storage methods are multiply, data can't be extracted directly through a unified data interface.

(2) Data distributed storage problem. Signal data around is distributed storage independently, which makes it difficult to ETL tool for flexible on-demand data extraction operation and concentration data processing.

For multi-format data sources and distributed storage problems, developing multiple data sources interfaces and daemons solution are adopted, which are shown as Figure 1 and 2.

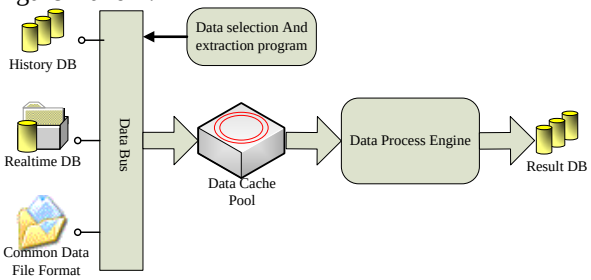

Figure.1 Data extraction method of multiple data sources interfaces

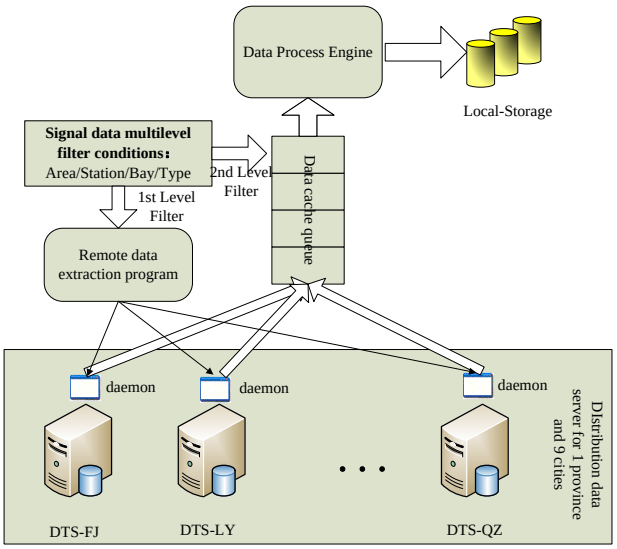

Figure.2 Daemon data extraction method for distributed storage

As shown in Figure 1, for real-time DB, history DB, CIME files and other data sources, data access interfaces were developed to constitute data bus. Users will choose a data source via the data selection and extraction program that calls the data bus to access the data source interface and extracts data to the data buffer pool in local memory. Data processing engine acquires stream data from the buffer pool for standardization and stores the result into the local results database.

As shown in Figure 2, for the data around stored independently, daemons are developed to implant into data servers around. Daemon starts with the operating system, memory resident, whose only task are to receive remote commands and call the local data access interface. Data extraction conditions are divided into two levels of filters. Remote data extraction program sent obtain instructions to each server daemon according to the first-level condition, that regional and plant station conditions. Daemon calls local data access interface in data server and transfers data to the data buffer queue. Data processing engine obtain data from the cache queue for processing according to the second-level condition, that the bay and signal type conditions. This will not only improve the efficiency of data extraction, but also prevent the data buffer queue be too large, while ensuring the system time complexity and space complexity.

OPERATION AND MONITORING SIGNAL BIG DATA CLEANING AND TRANSFORMATION ENGINE

CEP engine based operation and monitoring signal big data processing architecture

Due to the CEP engine has a good effect for complex transactional data processing, it has been used for standardization cleaning of operation and monitoring signal data. Figure 3 shows this architecture.

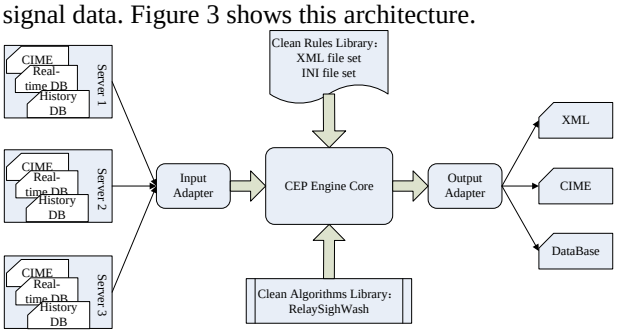

Figure 3 CEP Engine Architecture of Operation and Monitoring Signal Data

Operation and monitoring signal data processing engine architecture mainly includes the input and output module, the CEP engine core, the rules library and algorithms library. According to the character of signal data storage, the input adapter extracts each format source data in different servers to engine. The core of the engine is a container, which can provide input and output interface, signal processing framework, rule library, algorithms library and other functions. The format of signal cleaning rule library is XML file or INI configure file, which is mainly used to define cleaning rules such as keyword extraction, text replacement and text elimination in signal standardization process. The algorithms library is integrated in RelaySighWash class, which is mainly used to define and implement the intelligent algorithm interfaces such as algorithm of device and signal cleaning, word frequency analysis, pattern matching and so on. The rule library and algorithms library are the hard core of processing engine, which are designed to be pluggable in order to be easy to maintain and expand.

Data cleaning rule library

Signal data cleaning rule library is used to define the various rules according by the processing logic standardization after the data stream flow into CEP engine, which should have generality, intuition, easy to read, easy to maintenance and other features.

Device cleaning rule library

Cleaning signal first need to cleaning the protection devices' names signal belong. All rules of devices cleaning are defined into an XML file including extraction rule of basic device name and matching rule of typical device type. The file format is as follows:

```
<?xml version="1.0" encoding="GB2312"?>
<root>
<devclean>
CSC;CSR;ISA;CSI;PCS;WCB;WCH;UDL;FCK;DF;CAS;iPACS;CSL
</devclean>
<character>
a-zA-Z0-9/
```

</character>
<excess>
\\s-_
</excess>
</root>

Device name cleaning rules contain root tag “root” and two kinds of rule tags: “devclean” and “devname”. Names of all basic devices are defined as some keywords in the “devclean” tag, and thus all irrelevant contents in front of the keywords would be cleaned. For example, “non electric power RCS9661” would be cleaned as “RCS9661” by using the “devclean” rule, which extracts “RCS” and delete irrelevant information “non electric power”. “Character” tag is used to define all matched devices models, in which regular expression definition method is used to represent numbers, forward slash, upper and lower case letters. “Excess” tag is used to define other redundant characters, such as line and all kinds of spaces. In the same way, regular expression definition method is applied in this tag, in which “\\s” represents blank space, such as tabs and multiple spaces, and double backslash represents escape character. In accordance with this rules, the original device name of "32 / ARCS - 9705 - c measurement" can be intelligently processed as a specification name of "ARCS9705C". The details of processing algorithms are illustrated in section 2.3.

Signal cleaning rule library

Device operation and monitoring signal’s standardization cleaning rule library includes replacement rule of key-word in signal, elimination rule of device name in signal, pattern recognition rule of signal, whose format is define as follows:

<?xml version="1.0" encoding="GB2312"?>
<root>
<delete>
第一套;第二套;第Ⅰ套;第Ⅱ套
</delete>
<replace>
<key>Ⅰ段</key>
<value>I 段;一段;1 段;壹段</value>
</replace>
<combine>
<result>中后备动作</result>
<mode>中;后备;动作</mode>
<mode>中;后备;电笛;位</mode>
<mode>中后备</mode>
</combine>
</root>

Signal cleaning process can be divided into three processes, which defines cleaning rules according to delete tag, replace tag and combine tag, respectively. The delete tag is applied to define deletion rules, which deletes irrelevant entries in the signal. The replace rules are determined by the “value” and “key” child tags in replace tag. Keywords need to be replaced are deposited into the value tag with semicolons, and the replace results are deposited into key tag. The combine tag is used to define matching rules, which contains of “mode” and “result” child tags. The mode tag stores models, namely keywords combinations with the same meaning. On the other hand, the result tag stores standard signal expressions, namely results matching each model.

Data cleaning algorithms library

Signal data cleaning algorithms library includes many signal cleaning methods, which not only can be called by CEP engine but also can be called independently.

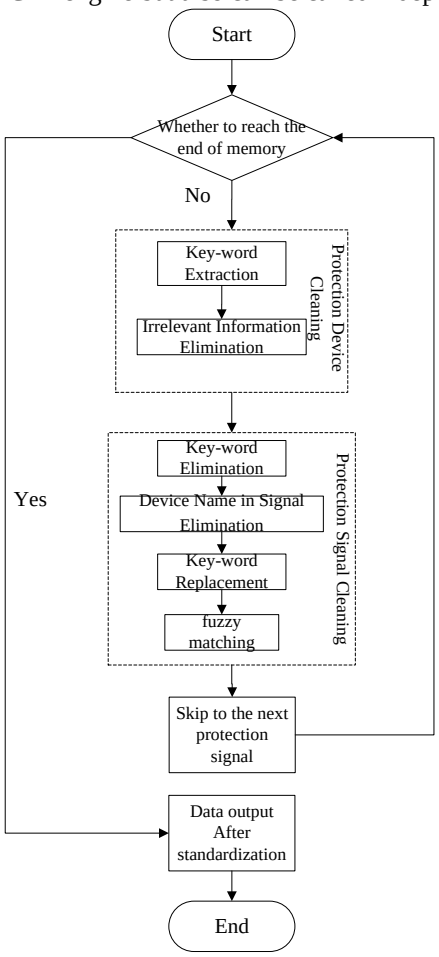

Figure 4 Flow chart of signal data cleaning algorithm

The algorithm mainly includes protection device cleaning and protection signal cleaning. The part of protection device cleaning includes key-word extraction and irrelevant information elimination in protector name two steps. Protection signal cleaning includes key-word extraction and elimination, protector name elimination in signal, key-word replacement and fuzzy matching and intelligent generated entry four steps.

SUMMARY AND SCOPE

The paper mainly introduced a CEP based big data ETL method for active distribution network operation monitoring and controlling signal. The multi-sourced and distributed signal on-demand extraction solution is proposed by analyzing current signal data storage. The pluggable cleaning rules library and signal library in CEP engine are formulated according to the signal features. In the end, the operation and monitoring signal load mode is designed according to electric system simulation demand, so that to drive simulation logic operation. The CEP engine based monitoring and controlling signal data processing method can implement fast and accurate cleaning of signal, so that providing effective data service for each signal driven power system applications.

The current operation and monitoring signal cleaning rule library uses more of enumeration mode, less machine learning method, which is not conducive to the maintenance of the rule library. So self-learning of the signal cleaning rule library is an improvement direction in the future.

ACKNOWLEDGEMENTS

This paper is supported by 2016 Beijing Natural Science Foundation (No.3161002) and Science and Technology Project of State Grid Corporation of China, *Energy Internet Oriented Application Researches of Multi-source Coordination Optimization Operation* and the State Grid Corporation of China headquarters' science and technology project *Key technologies research and pilot applications of Cloud computing technology in areas of scheduling model management, simulation training, protection setting and plan checking*.

REFERENCES

[1] Wang Jinli, Sheng Wanxing, Wang Jinyu, etc. Design and Implementation of a Centralized Data Acquisition and Supervisory System for Medium-low Voltage Distribution[J]. *Automation of Electric Power Systems*, 2012,18:72-76+81.

[2] Duan Cheng, Wang Zengping, Wu Kehe. Design and implementation of a lightweight ETL system for power real-time data[J]. *POWER SYSTEM PROTECTION AND CONTROL*, 2010,38(18):174-177.

[3] Meng You, Luan Zhong-Zhi, Xie Ming, Operator-Based Extendable Complex Event Processing Model[J]. *Journal of Software*, 2014,(11):2715-2730.

[4] Deng Huifang, Huang Juncheng. System Integration and Data Exchange Solution Based on Message Queue Software Bus[J]. *Journal of south china university of technology*, 2008,36(5):1-5.

[5] Yuan Hao, Qu Gang, Zhuang Weijin, etc. Discussion on Condition Monitoring Contents of Secondary Equipment in Power Grid[J]. *Automation of Electric Power Systems*, 2014,(12):100-106.