Fusion Engineering and Design 129 (2018) 88–93

Contents lists available at ScienceDirect

Fusion Engineering and Design

journal homepage: www.elsevier.com/locate/fusengdes

Study of web-based management for EAST MDSplus data system

Feng Wang[a,*], Yueting Wang[a,b], Ying Chen[a], Shi Li[a], Fei Yang[a,c]

[a] Institute of Plasma Physics, Chinese Academy of Sciences, Hefei, Anhui, China
[b] School of Nuclear Science and Technology, University of Science and Technology of China, Hefei, Anhui, China
[c] Department of Computer Science, Anhui Medical University, Hefei, Anhui, China

ARTICLE INFO

Keywords:
East
MDSplus
Data management
Data statistics
Data retrieval

ABSTRACT

A large distributed data acquisition system based on PXI/PXIe technology has been implemented on EAST tokamak. Now the whole data system has more than 60 data acquisition units and more than 3000 raw channels, and the maximum data throughput is about 5GBytes/s. Most of the acquired data are continuously stored into MDSplus Wan et al. database, moreover the calculated and analyzed data are also saved into MDSplus. In case of long-term discharge mode, the raw data of one shot is more than 1TBytes, and the total data of one campaign is more than 100TBytes.

MDSplus is helpful to store tokamak experiment discharge data, and the scientific data are saved into database by shot and tree, and it is easy to know the signal information of one shot or tree. However, it is difficult to get the general metadata information of all the shots and trees. Even if we can read the information by some scripts, but it is not so convenient and the data retrieval speed is slow. So we have planned to design a data statistics and retrieval system for EAST MDSplus data system, which can provide all the metadata information of all the trees such as signal name, sampling rate, data size, data available or not, and so on, basing on these metadata the system can give detailed statistics and reports automatically by some pre-defined model, which can help users to grasp the status and schema of all the data. The system has been developed based on apache/mysql/php framework and the PCIe SSD storage is adopted to promote the database access speed. The design details will be given in the paper.

1. Introduction

For the purpose of continuous data acquisition, a large distributed data acquisition system based on PXI/PXIe technology has been implemented. Now the whole data system has more than 60 data acquisition units and more than 3000 raw channels, and the maximum data throughput is about 5GBytes/s. Most of the acquired data are continuously stored into MDSplus database, moreover the calculated and analyzed data are also saved into MDSplus. In case of long-term discharge mode, the raw data of one shot is more than 1TBytes, and the total data of one campaign is more than 100TBytes [1–4].

MDSplus is helpful to store tokamak experiment discharge data, and the scientific data are saved into database by shot and tree, and it is easy to know the signal information of one shot or tree. However, it is difficult to get the general metadata information of all the shots and trees [5–8]. Even if we can read the information by some scripts, but it is not so convenient [9,10].

So we want to design a web-based management to get all data information including data status monitoring, tree files status, signal data retrieval, data statistics and report to help administrator to manage data.

2. System design

2.1. Architecture

The architecture of data management is shown in Fig. 1.

The system is designed based on standard application framework of Linux/Apache/MySQL/PHP (LAMP).

1) The metadata of all the MDSplus trees are extracted and saved into MySQL relational database;
2) Some of the key-value data are also cached to Redis NoSQL database to improve access speed;
3) Users can access the system using web browser after authentication;
4) The user account can be integrated with local LDAP account system;
5) The data management service is developed by GNU C/C++ language and it can communicated with the EAST Central Control

* Corresponding author.
E-mail address: wangfeng@ipp.ac.cn (F. Wang).

https://doi.org/10.1016/j.fusengdes.2018.02.068
Received 23 June 2017; Received in revised form 16 February 2018; Accepted 17 February 2018

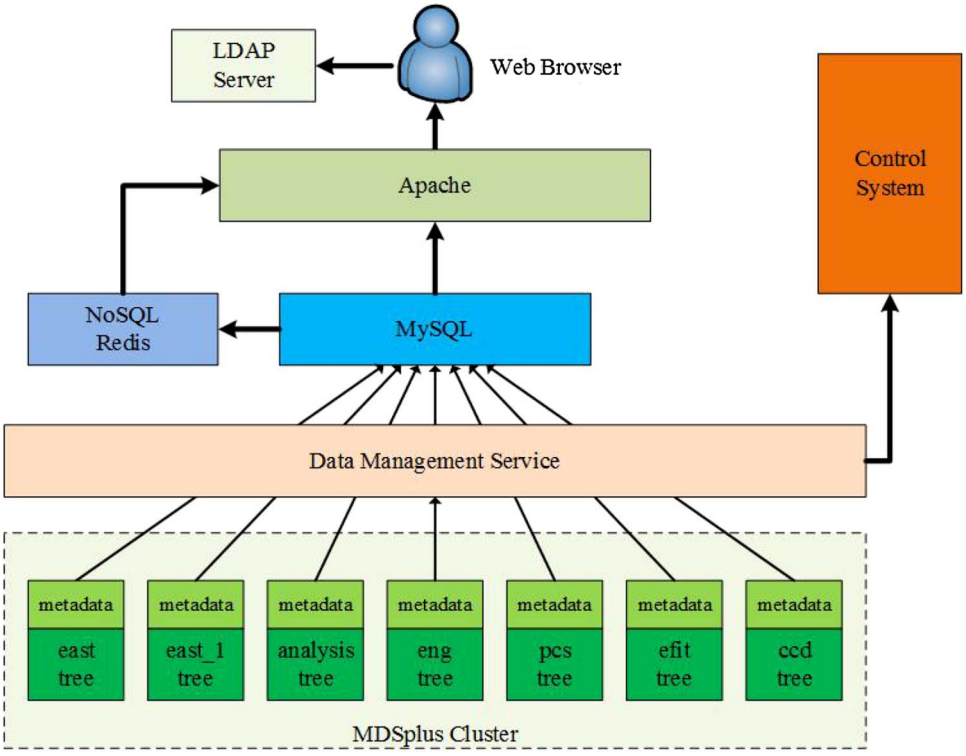

Fig. 1. System architecture of data management.

System;

6) The system can run in two modes, auto mode and manual mode. In case of auto mode, it can update the data automatically after each shot finished during discharge. Administrator can also execute the update function by manual.

2.2. Database design

The MDSplus software is designed as tree structure to store the complex scientific data. However tree structure is not easy to save and query in MySQL database. In order to simplify the problem, we only extract the important metadata related with signals including raw and analyzed data as shown in Table 1. The metadata of each tree are saved into one table and all the tables are listed below as shown in Table 2.

The latest shot number of last EAST campaign is 71818 and the total record number is about 700 million, and it will soon reach at 1 billion in the near future. Then it becomes a big problem to query in the large amount of records, so we have to consider about database query optimization (Table 3).

Table 1
MySQL table structure.

Field name	Data type	Sample data
shot	int	50000
subtree	varchar	t1
signame	varchar	ipm
createtime	timestamp	"2016/03/17 10:21:32"
freq	float	10000LU
ftrans	float	−184.212
fzero	float	0
fzoom	float	1
sigunit	varchar	"KA"
trigtime	int	−7
uimax	int	65536LU
uiswing	int	20LU
samplenum	int	150000

Table 2
Record number of tables.

Tree name	Table name	Record number (million)
east	mds_east	~210
east_1	mds_east_1	~210
analysis	mds_analysis	~140
eng	mds_eng	~140
pcs	mds_pcs	~35
efit	mds_efit	~7
ccd	mds_ccd	~1
	Total	~740

Table 3
Test results of query performance.

Test case	east (ms)	analysis (ms)	eng (ms)	pcs (ms)
Query all the signal names of one shot;	7.4	11.7	6.3	12.5
Query all the signal names of one shot;	1.4	1.0	0.8	0.4
Query all the shots having some signal;	91.7	81	92	85
Query the signal name of the latest 100 shots;	0.82	0.94	0.66	0.84
Query the signal data size of the latest 100 shots;	1.4	1.5	0.8	1.75

2.3. Query optimization

The most effective method for query optimization is to design reasonable indexes, so we have designed and adopted different kinds of indexes in the database.

1) Single index (idx_shot);
2) Composite index (idx1: idx_shot, idx_signame, idx_subtree);
3) Covering indexes (idx2: idx_signame, idx_shot, idx_subtree, samplenum);

After creating and using these indexes, the query speed promote about 100 times than that without indexes.