

Proficiency Testing Statistics Analysis System Based on Cloud Computing

Wu Gang¹, Zhao Chao¹, Zhao Jing¹, Zhang Fan¹, Wang Haitao¹, Cao Xinyu^{1,*}

¹China National Institute of Standardization, Beijing, China

*Corresponding author

Email: caoxy@cnis.gov.cn

Abstract. Cloud computing technology enabled to solve big data problems like data storage, process and data security. This paper proposed a design and implement method by cloud computing technology to build a system for proficiency testing statistics analysis. In proficiency testing system, analysis massive data and calculate a laboratory evaluation result is the core system function, through the cloud platform, integrated cloud platform resources to implement data distributed storage, numerical calculation and statistics analysis. The experimental result shows that the system deploy cost reduced than before and user experience obviously improvement on speed of system access and operation responses.

1. Introduction

Nowadays, cloud computing applied in more and more peers and enterprises. As the increase of data, traditional web system deploys on single or more than one server doesn't load vast amount of user's access and data process requirements. Modern laboratory manager refers to advanced information technology to improve efficiency, reduce cost, and normalize laboratory manager system. Proficiency testing(PT) aim to use the proficiency testing of interlaboratory comparison to confirm the implementation ability of special testing or measurement of laboratory, and monitor persistent executive ability of them [1].

The activities of proficiency testing include proficiency testing scheme, interlaboratory comparison and measurement audit. Proficiency testing provider use statistics methods[2] to evaluate the capability of laboratories, this paper to solve apply the information technologies based on cloud to implement these statistics methods. The statistics methods will introduce in section 3.4. As the data statistics analysis refer to the big data process, apply cloud computer and cluster multiple servers to implement parallel process has become the solution.

The organization of this paper is as follows: Section 2 presents the cloud computer technology. Section 3 introduces the system architecture and modules design. The implementation is explained in Section 4. The conclusion is given in section 5.

2. Key technologies of PT Statistics System

2.1. Kafka

Apache Kafka[3] is a distributed publish-subscribe messaging system and a robust queue that can handle a high volume of data and enables you to pass messages from one end-point to another. Kafka is suitable for both offline and online message consumption. Kafka messages are persisted on the disk



and replicated within the cluster to prevent data loss. It integrates very well with Apache Storm and Spark for real-time streaming data analysis. In the paper, we apply the Kafka to implement message subscribe for users.

2.2. MongoDB

MongoDB[4] is a free and open-source cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with schemas. MongoDB is developed by MongoDB Inc., and is published under a combination of the GNU Affero General Public License and the Apache License. In this system, we apply MongoDB storage the documents like testing report.

3. The Design of Proficiency Testing System

3.1. System Architecture Overview

PT business processes management and PT statistics analysis are critical functions in this system. PT business processes response to implement user management, PTP management, sample management, and testing result management, collect and manage testing data from PT participants eventually. The statistics analysis focus on assess the laboratory capacity through variously statistics methods to process and analysis testing data. In this paper, we design and implement a module named DataHub to collect PT data from participants and DataOLAP module to process and analysis data based on cloud platform, the system architecture is as the Figure 3.1 shows.

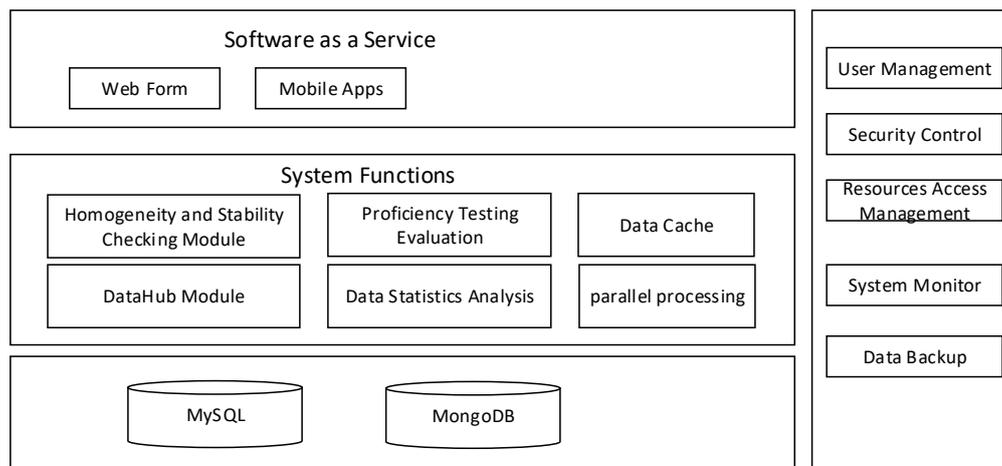


Figure 3.1 the diagram of system architecture

3.2. DataHub

Data manage module named DataHub in this system, it is main to collect and storage data uploaded from PT participants, the data includes structured and unstructured data. Structured data is testing result of PT scheme projects, one projects includes many participate laboratories, and one laboratories may submit more than one testing result, the testing result format usually is numerical or quantitative data. Except structured data like testing result data, participants also upload testing report which data as unstructured data, we apply MongoDB to storage it.

DataHub provide three interfaces for external application invoke, include data storage interface, data query interface and data subscribe interface, all the application call these 3 interfaces to implement data management function. The role of DataHub in the system is as the figure 3.2 shows.

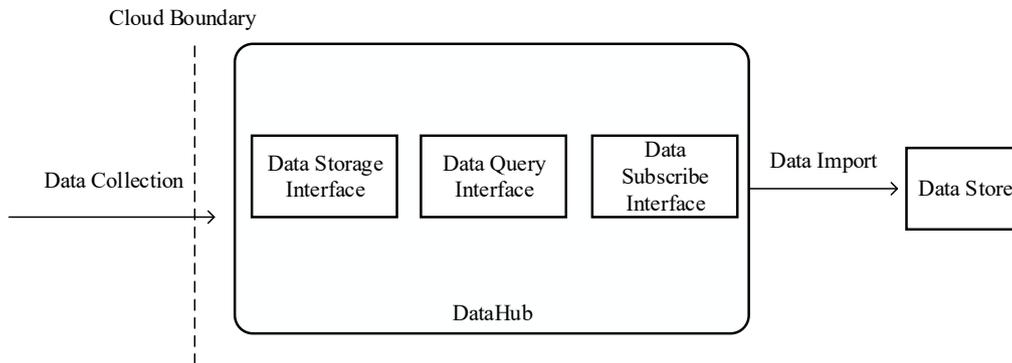


Figure 3.2 the role of DataHub in the system

3.3. Proficiency Testing Evaluation Module Design

Proficiency testing evaluation module main to implement proficiency testing scheme public, scheme apply and testing result statistics analysis functions. User input proficiency testing scheme information in the web form invoke DataHub module storage data to MySQL, laboratory users upload testing numerical data save to database, and save the testing report files to MongoDB. Invoke the statistics analysis module to implement testing data statistics analysis, eventually, represent the evaluate result by visualization technology. The process diagram of proficiency testing evaluation as the Figure 3.3 shows.

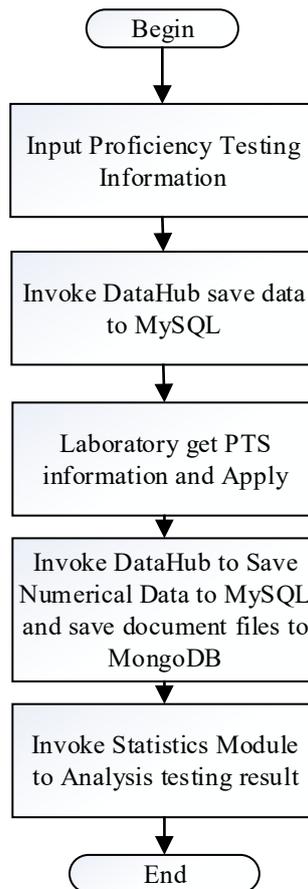


Figure 3.3 The process diagram of proficiency testing evaluation

3.4. DataOLAP Module

To implement laboratory proficiency testing evaluation, apply multiple statistics methods to analysis laboratory abilities is technological means. The statistics methods include confirm assign value and standard uncertainty values, confirm standard deviation of proficiency testing, calculate performance statistics and laboratory proficiency testing evaluation. In this paper, we designed 3 classes to implement the statistics methods. The class diagram as the figure 3-4 shows. The class ConfirmStdDeviation responses to implement confirm standard deviation, the class CapacityEvaluation responses to implement laboratory capacity evaluation. This module is isolate with DataHub, which invoke DataHub to import data for statistics analysis.

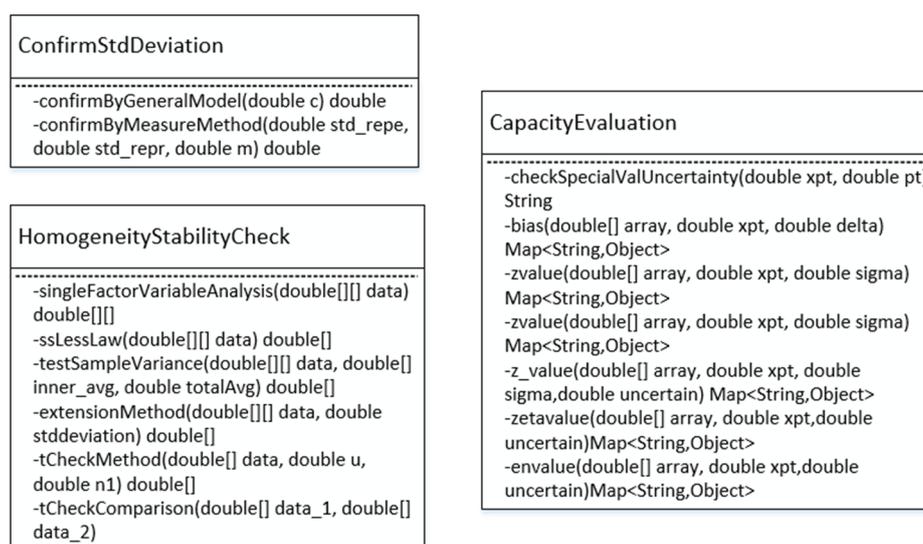


Figure 3-4 the Class Diagram of DataOLAP

4. Conclusion

In view the big data era, apply the cloud computer technology to implement business process system is present the general trend. This paper proposed a statistics analysis system to evaluate proficiency testing for laboratories, apply cloud computer technology, high efficient to implement data analysis and numeric calculation. The system design follows the loose coupling principle, by DataHub and DataOLAP module, shape a logic of data import by DataHub and data process and output by DataOLAP module. In DataOLAP module, implement all the statistics methods to evaluate the capacity for laboratories, be good for popularize the standards, improve the laboratory's capability compare with the international laboratory.

Acknowledgement

This paper is supported by grants from National Key R&D Program of China (2016YFF0204205, 2018YFF0213901) and China National Institute of Standardization (522016Y-4681, 522018Y-5948, 522018Y-5941).

References

- [1] GB/T 28043-2011 Statistics methods for use in proficiency testing by interlaboratory comparison
- [2] Shengnan Liu, Gang Wu, Chao Zhao, Haitao Wang, Jing Zhao, Fan Zhang. The comparison of three robust statistical methods in proficiency testing. ICCD 2017.
- [3] <http://kafka.apache.org/>

- [4] <https://www.mongodb.com/>
- [5] Penin, Philippe. Comparaisons interlaboratoires. 4th International Metrology Conference 2012, CAFMET 2012, p 52-56.
- [6] Li, Zhang. Application of a resource-sharing platform based on cloud computing technology. Agro Food Industry Hi-Tech, v 28, n 1, p 2205-2209, 2017.
- [7] Ahmed Khan, Muhammad Naeem; Ullah, ShahWali. A log aggregation forensic analysis framework for cloud computing environments. 1st International Conference on Big Data and Cloud Computing, ICoBiC 2017.