

Design of Test Data Management System Architecture Based on Cloud Computing Platform

Junbin Duan, Pengcheng Fu, An Gong, Zhengfan Zhao
Beijing Institute of Special Electromechanical Technology,
Beijing, 100012, China

Abstract—Test is the major means of verification and validation for a complex electromechanical system development process. Test data management(TDM) system unified management of various types of documents, structured data, and other test resources. In the course of the development of large scale system, it involved many types of data, large amount of data to terabytes. At present, most of the TDM used a relational database management system for storage, so there are difficult type expansion, retrieval and inefficient when dealing with large data. The TDM system architecture based on cloud computing platform adopts the distributed technology with ability to achieve high reliability and flexible expansibility. This article describes the system applied three layer architecture based on Hadoop, including cloud storage layer, service layer and application layer of the main composition and function, which work together to achieve the flexibility and expansibility. Based on this system architecture designed a TDM system, which realized the scalability of a variety of heterogeneous data management, and could be used for real time data query.

Keywords-cloud computing; test data management; system architecture; Hadoop.

I. INTRODUCTION

China is transforming from “Made in China” to “Created in China”. As technology contents of large machinery and electronic equipment increase continuously, systems become more and more complex. Lots of data are generated during tests, and various kinds of data are widely distributed. Data management problems become more and more prominent, and various kinds of design, simulation, calculation data and test data generated during development become key system contents. How to effectively manage these data becomes one of the key initiatives of integrating IT application with industrialization.

Today, cloud computing technology has become the hot research direction of major Internet enterprises. Commercial giants such as Amazon, Microsoft, Intel and Google, have successively launched a series of products and displayed their research results.[1] In order to solve many problems of existing information system including not fully open communication protocols, not strong functional expansibility, not high system scalability, and complexity of later update and maintenance, the basic cloud platform system architecture of test data management system is designed by using the advanced technologies such as

network communications and cloud computing, and by adopting SOA and cloud architecture system. The data management system architecture with high reliability, strong expansibility, high scalability and openness, is available.

II. STATUS QUO OF TEST DATA MANAGEMENT SYSTEM

With the further development, test quantity and test data will increase exponentially. If the manual management mode is used, testing personnel, due to large workload, heavy tasks and high proportion of manual calculation and interpretation, will encounter certain mistakes during the process of data detection and report treatment. As a result, automation degree of calculating, processing and judging test data is not high, and test quality can not be guaranteed better.

Situation of tests for development, information required for test management decision-making, and summary data accumulated, can not be analyzed effectively; accumulated data can not be retrieved fast, and situation of test detection can not be reflected.

Thus, it is necessary to save test data, test documents and test resources (which are scattered among components, tests and test equipments) in the unified data storage platform, unify entries and exits of test data, and make users able to conduct various operations through the unified data platform, so as to reduce the error rate of test data and improve the utilization efficiency of test data.

Test data are different from other data, and have unique characteristics:

i) Test data are composed of structured data and unstructured data, including image, figure, text, video, audio and other types of data.

ii) Complex data processing: Real-time processing, pre-processing and post-processing.

iii) Many measurement parameters: There are as many as 10,000 measurement parameters for R&D and test of one product.

iv) Some data records are synthesized by several kinds of data. Definition of data attributes such as temperature and pressure should be extended flexibly.

v) Tremendous test data. Data size of a single test module or the entire test is as high as dozens of G.

III. CLOUD COMPUTING TECHNOLOGY

Cloud computing is the inevitable result of developing technologies such as distributed processing system, parallel processing and grid computing to meet modern service needs. Cloud computing is a new computing model, and also a new mode of combining computer resources, representing more an innovative business model [2].

As a new computing architecture and application model, cloud computing has the following basic characteristics [3]: (1) High reliability. Relatively mature and widely used technologies such as distributed computing, virtualization and grid computing, guarantee the reliability of cloud computing technology; to ensure security of cloud computing, data distributed in different servers are subject to multiple copy fault-tolerance, and computational nodes are isomorphic and interchangeable; in order to survive and develop in the fierce market competition, the cloud platform and the secondary development based on the platform stand the test in quality and quantity. (2) Ultra-large scale and scalability. "Cloud" is boundless, and cloud computing is characterized by ultra-large scale in terms of infrastructure facilities, information base, information service scope and information users. Cloud computing can be seamlessly expanded to large-scale clusters, and even several thousand nodes can be processed simultaneously. Users think that "cloud" can be dynamically scalable to meet the service needs of different users in different periods.

From the perspective of user access mode of cloud computing, "cloud" can be divided into public cloud and private cloud. Private cloud provides the security of traditional preset infrastructure, and also scalability of relatively new cloud computing model. In contrast with public cloud, private cloud provides more reliable uptime and better tracking service, and private cloud users have absolute control over network. Meanwhile, it also provides better scalability than the preset architecture, because users can extend with increase of needs. To achieve the fundamental purpose of data privacy and service integrity, private cloud is the best choice of enterprises. Currently, Hadoop is a private cloud computing platform most widely used. Hadoop is a project product developed by the open-source community Apache by cloning the GFS thinking of Google, and is a distributed cloud computing system developed according to MapReduce and GFS international papers which are published by Google. Hadoop is a distributed file system for storage, and also a framework for distributed applications on large clusters composed of general computing devices. As part of Hadoop,

HBase can read and write the distributed big data (structured data or unstructured data) at a high speed. HBase serves as the data storage container of this system.

IV. TEST DATA MANAGEMENT SYSTEM ARCHITECTURE BASED ON HADOOP

Currently, most test data management systems save all system data through the method based on relational database. Structure of relational database is hard to change, and data have fixed length, thus limiting the flexibility of data model design. A lot of development work is needed for different types of product data, which restricts the applicability of test data management system. In addition, as data are accumulated continuously, retrieval and maintenance of mass data challenge the storage and processing capabilities of relational databases. As a result, the speed of data retrieval and processing can not meet business needs, and hardware often needs to be re-configured, upgraded and optimized. As the cloud computing technology becomes mature, the test data management system based on HBase database and HDFS file system in Hadoop provides a solution. HBase is a column-oriented distributed database rather than a relational database, and is designed to solve the limitation of relational database in processing the massive test data.

For large-scale and intensive data applications, and all types of database storage systems and cloud storage systems, data models under the cloud computing environment should be established according to theories and models of cloud computing and cloud storage, to provide the transparent and unified data integration and access interface services. This system architecture can realize the intelligent integration of various test data under the cloud computing environment, and can meet the requirement for high user concurrency, high load and high-speed processing of mass data.

The system architecture establishes SOA services according to the characteristics of cloud computing environment such as virtualization, distribution, high reliability and high scalability, and achieves data storage and management based on HBase and HDFS. The overall technical model of the system, from bottom to top, includes cloud data layer, service layer and application layer. The three layers operate under the unified management and coordination of cloud computing data system management module, and the overall logical framework of the system is shown in the figure below.

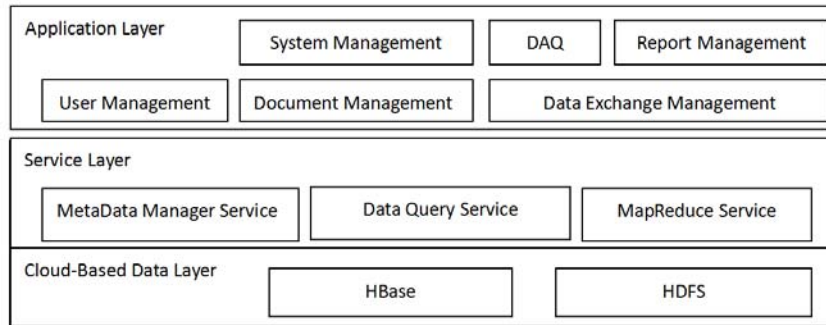


Figure 1. Logical framework of the system

A. Cloud Data Layer

Under this model, cloud data layer serves as a unified database, and is mainly used to store the structured data and documents. The two-dimensional table of traditional relational database is column-stored based on HBase. Table indexes include Row Keyword, Column Family, Column Keyword, and Timestamp. HBase table is four-dimensional. Column Family should be determined in advance when designing the table structure, and Row Keyword, Column Keyword, and Timestamp may be added dynamically. Each Column Family corresponds to a class of structured test data, and data columns can be increased or decreased dynamically. Through metadata management, users can dynamically modify data types, data column definitions and keywords, so as to describe, obtain and analyze requests more efficiently, accurately and conveniently, and obtain data quickly.

B. Service Layer

Service layer consists of metadata management service module, data inquiry service module and MapReduce service module. Inquiry service module accesses the data layer through MapReduce service module. Data maybe exist in many formats: multimedia, image, text, real-time data, sensor data, etc. New data sources maybe require new formats. MapReduce can store and analyze various raw data formats.

Metadata management service of this layer is used to create the user-defined data structure. The data sheet function provided by the Hadoop system is adopted, and corresponding column families and data columns are synchronously established with the metadata definition. Data such as documents and graphics are stored in the form of byte, and basic attribute data are stored in the data column.

C. Application Layer

Through the interface of service layer, application layer interacts with the user, and is mainly responsible for sending data requests, displaying data, collecting and integrating test data.

This layer is developed by using JSP. When data are collected, data edit page is generated according to metadata definition, and the user can input data. Through the DAQ

collection software, the user can send the collected test data to the data server via MapReduce, thus achieving automatic collection of test data. In consideration of different data storage models under the cloud computing environment, application layer must integrate existing resources. The unified data access interface must be provided to meet the requirements for system scalability.

V. CONCLUSION

This paper presents a kind of test data management system architecture based on cloud computing platform. This architecture is characterized by distinct gradation, easy expansibility, flexible data structure and reasonable overall design, and applies to data management of test businesses. At present, this system architecture based on Intel Hadoop has been applied for the test data management of large manufacturing equipment. The user can conduct the real-time inquiry, statistics and analysis of various kinds of test data distributed in cloud terminals, and generate various types of data analysis reports according to the report templates, which significantly improve the level of managing and using test data. The technologies adopted for this system have reference significance for other similar systems. Currently, application of cloud computing technology is still in its infancy, and many complex data applications need the further study.

REFERENCES

- [1] Intel Hadoop Release V2.2 Developer Guide, <http://www.intel.cn/content/www/cn/zh/big-data/intel-s-distribution-of-hadoop.html> [2013-1 - 28]
- [2] ZHANG Xiao-Juan, YI Min-Wei. System Integration Architecture and Enterprise Case Study in the Cloud Computing and SOA Context, Computer System Application, 2011 [9]:1-6
- [3] JIANG Wu-Xue, ZHANG Jing, WANG Zhiming. Cloud computing and its architecture model, Journal of Liaoning Technical University (Natural Science), 2011 [8], V01-30 No.4:576-579