

Design of Safety Production Control System Based on Big Data Cloud Computing Platform

Ganglong Fan¹ and Yongliang Li^{1, a*}

¹Luoyang Normal University, Henan Luoyang, 471934, China

^aliyongliangls@163.com

*The corresponding author

Keywords: Safety production control; Big data; Cloud platform; Server virtualization; Hadoop

Abstract. This article describes the big data analysis and processing methods based on cloud platform, and then this paper discusses the function of the safety production supervision system. Big data processing technology can effectively analyze the huge amount of data in the safety supervision, and provide intelligent service for managers. The paper presents design of safety production control system based on big data cloud computing platform. The validity of the proposed method is verified by experiments.

Introduction

Big data is a product of the information age, although in recent years, our country in the integration of the two has done a lot of work to promote safe production, production safety and other aspects of information technology, has made great progress, but overall China's safety production normalization level is still relatively low, little more than collecting applications, light weight later in advance such prominent problems, obstacles the application of large data. Second, the lack of high performance and large data analysis tools, which is widely, used in various fields of application of large data problems, if there is no high performance analysis tools, the value of big data can not be released. Third, since the promulgation and implementation of the government information disclosure regulations, the public security work has made a great breakthrough.

Safety is the safety supervision, fire protection, public security, transportation, special equipment, housing, environmental protection and other government departments, at the same time, the safety production data itself has a certain randomness in a short period of time, more comprehensive data for decision analysis can provide more valuable data [1]. In the case of ensuring data security, through data exchange, and it is sharing and other ways to break the barriers of data policy barriers, the integration of the various units of business data.

In the big data era, government data is the capital, the use of big data can not only realize the macro management of the tourism industry, but also to grasp the initiative of public service. For the government itself, to improve the level of e-government, e-government construction cluster system; strengthen the cooperation with traffic, weather, customs, public security departments at the same time, the typical internet tourism enterprises, tourism investment, individual consumption into the statistics system, to achieve cross sect oral, cross industry, cross regional the sharing of resources; establish tourism data exchange platform, the formation of data exchange and sharing mechanism.

Cloud computing environment to build the process can make full use of the relevant storage devices, and the introduction of the storage pool can be generated by the server, such as IBM or HP. The general selection of cloud storage devices to achieve data storage function is mainly due to the fast file storage and storage, etc.. In order to provide quality services to the environment, it is required to use private and total two kinds of storage cloud; the two are in charge of the internal and external services. The paper presents design of safety production control system based on big data cloud computing platform.

Construction of Big Data Processing System Based on Cloud Platform

Data dependence analysis is one of the most important supports of big data applications. Unlike the small data era, hybrid big data compels us to analysis in the unconditional distribution of data a priori hypotheses, data dependence analysis can not use subjective assumptions, observational sampling, statistical analysis, obtained the conclusion ", analysis method and more is the use of exploratory analysis method, the trend of mining association rules from an earlier start. The related data is divided into two aspects; one is the correlation between the features (variables), typically, such as one or multiple regression analysis [2].

An important purpose of data mining is the user to understand the data, but also is the understanding of multi granularity and multi perspective, for scientific data modeling, which is particularly important, in a certain sense determines the practicability of mining algorithm; automatic modeling algorithm is domain experts understand the data reasoning tools and auxiliary means. So the development of user friendly, and even the people in the loop (man-in-loop) reasoning model is very important.

Distributed unstructured parallel storage system is the massive information processing environment under the most ideal storage solutions, from architecture design is a good solution to the capacity expansion and the performance of the storage system expansion problem. Distributed unstructured parallel storage system using new technology to replace the original copy of the RAID technology, not only to ensure the safety of the data, but also improves the efficiency of data recovery, reliability, performance and maintainability in combination. Non structured parallel storage system is mainly composed of the index server clusters and storage cluster server, the software components of the application for a file system; non structured parallel storage system architecture adopts the design concept separating the mainstream control path and data path, as is shown by equation (1).

$$f(x) \xrightarrow{\text{CWT}} W_f(a,b), \quad f(x-x_0) \xrightarrow{\text{CWT}} W_f(a,b-x_0) \quad (1)$$

Server virtualization technology is an important means to achieve the pool of computing resources. Virtualization is an abstraction layer, which separates the physical hardware from the operating system, thus providing a higher IT resource utilization and flexibility [3]. Virtualization allows multiple virtual machines with different operating systems to run in parallel on the same physical machine. Each virtual machine has its own set of virtual hardware (such as CPU, RAM, network card, etc.), you can load the operating system and application in these hardware. Regardless of the actual physical hardware components, the operating system is regarded as a set of consistent and standardized hardware.

An instance of the system on data stream processing technology and popular are analyzed. The realization technology of data stream processing system concern here, the distributed stream processing system as the memory data management system for real-time processing are described. However, the current distributed data stream processing system needs the user code to realize the query operation of the data, and the system itself does not provide the corresponding query language. Therefore, in order to improve the processing system easy to use and processing ability, enhance the application reusability and portability, as is shown by equation (2), w should realize the relationship with abstract query language query system in distributed data processing system, to build a complete flow management system.

$$\Sigma_N = \begin{bmatrix} \lambda_{N+1} & & & \\ & \lambda_{N+2} & & \\ & & \ddots & \\ & & & \lambda_M \end{bmatrix} \quad (2)$$

Intelligent search engine is realized on large data management platform based on the public security organs, big data platform management framework of the public security organs, including

data source, data management, WEB service support, should be the data source from users upload documents and server storage resources and cyber source, big data management platform for big data upload the document through the reverse order segmentation stored in a distributed file system, application layer through the WEB browser to achieve the corresponding service query etc.. Key technologies of large data management for data file segmentation, storage, management and fast browsing [4].

Distributed storage not only effectively solve the problem of the performance of the storage system, greatly reduces the cost of large-scale storage systems, will also enhance the reliability of storage system design level; at the same time, the physical storage device and to store the actual logical representation of separated, realize the storage virtualization in the. Because of these characteristics of distributed storage, as is shown by equation (3), where it has been recognized by large system customers, and gradually become the mainstream of the current storage architecture of cloud computing platform.

$$\hat{X} = (H^T R^{-1} H)^{-1} H^T R^{-1} Z \quad (3)$$

In the data sets presented with some uncertainty function; the other is a sample (object) related, such as association rule, it is emphasized that the dependence of the collocation between objects with some property sets [5]. Traditionally, describe in statistics Pearson correlation coefficient is two. The random variable is function dependence (as with function curve of random disturbance on large data). In terms of the relationship does not only mean the function, it may also be of some value dependencies, distribution similarity, or characteristic with various uncertainty of collocation.

With the progress of the times, the data also changes, with a variety of complex forms. A lot of research institutions and individuals at the same time of the data mining in structured data also launched a data mining and analysis of spatial data and multimedia data, and time series data, text data and Web data. At the same time, the development of big data promotes the generation of cloud computing, and data mining based on cloud computing is rapidly rising.

Design of Safety Production Control System Based on Big Data Cloud Computing Platform

Big data, also known as massive data, is with the rapid development of computer technology and Internet technology and the unique data generated phenomenon [6]. The modern society is producing large amounts of data, such as web access, WeChat and micro-blog, video images, mobile phone communications, and online shopping, at unimaginable speeds..... And so on are constantly generating a lot of data.

Safety production responsibility system is a part of enterprise post responsibility system, and it is the most basic security system, and also is the core of enterprise safety production and labor protection management system. Practice has proved that, to establish and improve the safe production responsibility system in the enterprise, leaders attach importance to production safety, labor protection work, conscientiously implement the party's safety production, labor protection policy and national safety production, labor protection laws and regulations, the responsible organization of production at the same time, take active measures to improve the working conditions that will reduce accidents and occupation disease [7]. On the contrary, it is unclear, but the mutual prevarication, production safety, labor protection work can not be carried out, no one is responsible for industrial accidents and occupation disease will continue to occur.

The server virtualization technology can not only reduce capital consumption, reduce operating costs, so that the server can adapt and reconfigure fast, frequent, effective reduction in the development, testing, preparation and deployment cycle in time consumption, and rapid deployment, migration, allocation of resources in a reasonable strategy, realize resources as needed. Distribution, as is shown by equation (4).

$$\tilde{g}(n) = (-1)^{n-1} h(-n+1), g(n) = (-1)^{n-1} \tilde{h}(-n+1) \quad (4)$$

The data center through a unified enterprise security support platform to build a unified generation information registration function module based on the required uniform reporting and data reporting, unified unified filing supervision system implementation of all safety supervision departments at all levels, to avoid long enterprise data reporting and reporting, leading to inconsistent problems such as duplication of enterprise and government business data.

Collect the basic information related to the enterprise production safety, and provide a complete filing filing and approval mechanism [8]. The main system of hazardous chemicals enterprises and industrial enterprises, the mining enterprise basic information, organization information, security personnel, is mainly responsible for special operations personnel casualties, equipment and facilities and operation site, enterprise's qualification certificate, safety evaluation, labor supplies distribution situation, production safety system of census registration and filing management.

Big data fit the data needs to understand, it's easy to understand the information of particle (basic concepts) as the calculating unit and reasoning to atomic concept model as the basic representation tool to realize the multi granularity information extraction and inference, this computing paradigm easily people's prior information into the model structure, design in solution mechanism modeling, human-machine collaborative loop and complex problem [9]. Multi granularity reasoning human-machine cooperation will provide fast and efficient strategy to solve complex decision tasks under the big data environment also provides a flexible mechanism and algorithm of multi granularity understanding of big data and multi granularity reasoning for different levels of decision makers.

Experiments and Analysis

The safety data platform built safety production informatization based unified support platform (including data modeling engine, business process engine, report engine, engine system privileges etc.), to achieve a unified data standard and unified technical architecture, to achieve the city's informationization construction of production safety, the top-level design and construction, which can effectively avoid duplication of investment, departments at all levels the system is not compatible, data sharing the drawbacks of traditional construction mode.

In the face of complex tasks, interactive query and online processing, Hadoop and MapReduce are not applicable. Spark is a general-purpose parallel computing framework, developed by Berkeley AMP of the university laboratory, after Hadoop Spark has become a popular open source project; there are companies such as Intel added to the open source project [10]. This type of Spark/Shark memory computing framework is more suitable for the analysis of various iterative algorithms and interactive data, each time will be flexible distributed data sets (RDD) after the operation result is stored in the memory, the next operation can be read directly from the memory, as is shown by equation (5), where m can save a lot of disk IO will significantly enhance the efficiency.

$$c_k^j = \sum_{n=0}^{L-1} h(n)c_{2k+n}^{j-1}, \quad d_k^j = \sum_{n=0}^{L-1} g(n)c_{2k+n}^{j-1}, \quad (j = 1, 2, \dots, J, k = 0, 1, \dots, N/2^j - 1) \quad (5)$$

Let m be the number of samples is assumed, for the 1NN algorithm, the algorithm has higher classification efficiency, but when the concentration distribution of training samples, it is likely to lead to lower classification accuracy; and m=1, is the classic KNN algorithm. Therefore, the common k nearest neighbor algorithm is a special case of the large data processing algorithm proposed in this paper. Through the above analysis, we can know that the number of M cluster is larger, as is shown by equation (6) and the number of samples to be scanned by KNN is less, and the running speed is faster.

$$f(z) = p^H(z)U_N U_N^H p(z) \quad (6)$$

Variety provides us with the possibility to transform the seemingly useless information into useful information in the mass of data. Velocity can accelerate the data exchange, to meet the real-time needs of customers faster. Veracity is an important prerequisite for the application of large data; data analysis should be based on accurate basic data to make meaningful decisions. Value refers to the vast amount of data in a very small amount of data is valuable, we do is the use of cloud computing,

intelligent technology platforms such as mining valuable information, to facilitate the correct decision-making and action.

Although HDFS has become a big data recognized storage and MapReduce as its mix of data processing framework in the early stage of the development of large data show a significant value. Because of the design of the MapReduce constraint is only suitable for off-line calculation, there is still considerable problem in real time, some problems such as Hadoop iterative calculation, each time for the disk and network overhead is quite large. In particular, the calculation for each iteration results will be written to disk read back, in addition to the computation of the intermediate result also need three backup, which is a waste.

Conclusions

The application system of the safe production platform, emergency management, emergency management, emergency rescue and a series of links, whether in daily operation or rapid response, all need a unified process is to control. By the use of workflow technology can facilitate the definition of the system control process. The paper presents design of safety production control system based on big data cloud computing platform. The system can be put into automatic operation quickly, the process mechanism and response platform integrated application system of timely treatment.

Acknowledgements

This paper is supported by the science and technology research major project of Henan province Education Department (13B520155, 17B520026) and Henan Province basic and frontier technology research project (142300410303).

References

- [1] Yoon-Su Jeong, Kun-Hee Han, "Group Management Scheme for Security Assurance of Big Data in Cloud Environment", *JNIT*, Vol. 4, No. 8, pp. 561 ~ 568, 2013.
- [2] Woong Jang, Hong Gun Nam, Mu Wook Pyeon, Karp Joo Jeong, "New Trend Analysis of Spatial Information using Big Data Technology", *JNIT*, Vol. 4, No. 8, pp. 1 ~ 6, 2013.
- [3] Kyoo-Sung Noh, Seung-Hee Lee, "A Study on Innovative Model for Communication System of Political Parties in Korea by using Big data", *IJACT*, Vol. 5, No. 13, pp. 342 ~ 348, 2013.
- [4] P. Pradeepa, M. Jaiganesh, A. Vincent Antony Kumar, M. Karthiha Devi, "Fuzzy Based Reactive Resource Pricing in Cloud Computing", *JCIT*, Vol. 9, No. 6, pp. 130 ~ 142, 2014.
- [5] Hyung Woo Park, Heejun Yoon, Haengjin Jang, Jongsuk Ruth Lee, "Study on the Automatic Alteration of the Transmission Style for Big Data Transfer according to the Change of Internet Status", *JNIT*, Vol. 4, No. 9, pp. 89 ~ 95, 2013.
- [6] Taeho Kim, Gaewon Kim, Dai-Hwan Lim, Changhoon Yim, "U-Home Network Architecture Based on Cloud Computing", *RNIS*, Volume 12, pp. 161 ~ 164, 2013.
- [7] Rodrigo Rocha Silva, Celso Massaki Hirata, Joubert de Castro Lima, "Computing BIG data cubes with hybrid memory", *JCIT*, Vol. 11, No. 1, pp. 13 ~ 30, 2016.
- [8] YANG Zhexi, XUE Huacheng, "Strategy Study on Infrastructure Construction with Cloud Computing for Enterprises", *IJACT*, Vol. 5, No. 3, pp. 282 ~ 289, 2013.
- [9] Youngmahn Han, Hyunsik Kim, "Workflow-based Automation of Next-Generation Sequencing Data Analysis using Cloud Computing Resources", *JNIT*, Vol. 4, No. 9, pp. 109 ~ 113, 2013.
- [10] Zhou Min, "Analysis of Security Issues in Cloud Computing", *AISS*, Vol. 5, No. 5, pp. 493 ~ 500, 2013.