Building the Cloud Platform for the Next Generation Public Security Application

Xin Wang, Jie Dai and Zheng Xu
The Third Research Institute of the Ministry of Public Security, Shanghai, China

Keywords: Big Data, Public Security Information Systems, Cloud Platform, Distributed Storage, Distributed Computing, Data Retrieval, Virtualization.

Abstract: A great variety of public security information systems have been built for the traffic accidents governance, crimes events and terrorist incidents prediction. However, the large-scale redundant construction of systems leads to "great waste of IT resource" and "information overload. Technologies such as big data, cloud computing and virtualization have been applied in the public security industry to solve the above problems. This paper concludes a novel architecture for next generation public security system, and the "front + back" pattern is adopted. Under the architecture, cloud computing technologies such as distributed storage and computing, data retrieval of huge and heterogeneous data are introduced, and multiple optimized strategies to enhance the utilization of resources and efficiency of tasks.

1 INTRODUCTION

In recent years, a great variety of public security information systems have been built, which have played important roles in the traffic accidents governance, crimes events and terrorist incidents prediction. Series of problems appear, on the one hand, redundant construction of systems leads to great waste of resource, such as the video surveillance systems throughout the country, which are built with their independent software and hardware in each place. Furthermore, it is difficult to organize, manage and store the large-scale heterogeneous data including video, audio, text and structured data collected efficiently. And the most important is how to find valuable clues or knowledge quickly from great amount of information.

Technologies such as internet of things (Hu et al., 2014; Luo et al., 2011), big data (Xu et al., 2014; Xu et al., 2015) and cloud computing (Liu et al., 2010; Liu et al., 2011) have been applied in the public security industry to solve the above problems by governments all over the world. Utah Data Center, which was built for the American police and government, has the large-scale storage capacity of yotta bytes, and it has been collecting kinds of information, including personal e-mails, phone calls, parking receipts, travel schedules, shopping records and other records (www.nsa.gov). Boundless Informant project has developed a cloud platform which analyses data such as telephone, financial information and other intelligence transferred by wired and wireless network, satellite and other channels, a to achieve the global target of any real-time monitoring and network monitoring (www.nsa.gov1.info/dni/boundless-informant.html). ACCUMULO was a data storage software developed by the US National Security Agency, and submitted to Apache as an open source project in 2011 (www.accumulo.apache.org). Based on the Google's BigTable data model, structured and unstructured data are stored as distributed KV format, and the properties of database security, scalability and speed are enhanced. In Shandong province of China, the cloud platform for police was constructed in 2014 (www.prnasia.com), which provides applications such as "cloud search", "cloud video surveillance" and so on. There are 146 kinds of data acquired from kind of public security and other social information systems every day, and the number of structured data is up to 6 billion, and the platform has strong computing ability and store capacity for large-scale data.

Based on the research above this paper reviews the common architectures of the public security cloud platforms and data centers, and introduces the techniques for big data storage, organization, and analysis. The rest of the paper is arranged as follows. Section 2 introduces the problems. The architectures

* The corresponding author: Jie Dai

Wang X., Dai J. and Xu Z.
Building the Cloud Platform for the Next Generation Public Security Application.
DOI: 10.5220/0006020701420145
In *Proceedings of the Information Science and Management Engineering III (ISME 2015)*, pages 142-145
ISBN: 978-989-758-163-2

Building the Cloud Platform for the Next Generation Public Security Application

are detailed in section 3. In section 4, key techniques and optimized are described for the current problems faced by the public security of big data. Section 5 introduces the recent work and research. In the end, some conclusions are drawn and future work is given in section 6.

2 PROBLEMS DESCRIPTION

Redundant construction of information systems for public security governance, the traffic accidents, crimes events and terrorist incidents prediction leads to series of problems:

(1)Redundant construction of systems results in great waste of IT resource. There exist kinds of detective systems for police such as the video surveillance systems, which are built all over the country with independent software and hardware in each place. Redundant construction of systems leads to great waste of resource, it is still hard to discover deep information and complex content to support crime predictions without unified platform to integrate the distributed data.

(2)Data collected include video, audio, text and other unstructured, structured data, and it is necessary to store, manage, and process them under unified architecture and optimized strategies based on the cloud computing technologies, so that data computing and storage could be effectively.

3 THE ARCHITECTURE OF CLOUD PLATFORM FOR THE NEXT GENERATION PUBLIC SECURITY SYSTEM

We proposed the architecture of cloud platform for the next generation public security system, to show how to process, organize, manage and store large-scale heterogeneous data. As shown in Fig.1.

Due to the limited bandwidth, the “front + back” pattern is adopted, that is: data such as video, audio and other unstructured and structured data are collected by sensors such as cameras or from existed information systems and preprocessed in the “front” part, then the results are packaged with unified standard format and transferred to the “back” data center with strong storing and computing ability to support more complex computing and applications. The pattern could avoid network congestion caused by distributed heterogeneous data after data preprocessing in the “front”; in addition, based on the cloud computing and virtualization technologies, the cloud center realizes the resource consolidation of multiple IT resources, and provides unified computing and storage environment for more data analysis and applications such as data mining and semantic reasoning.

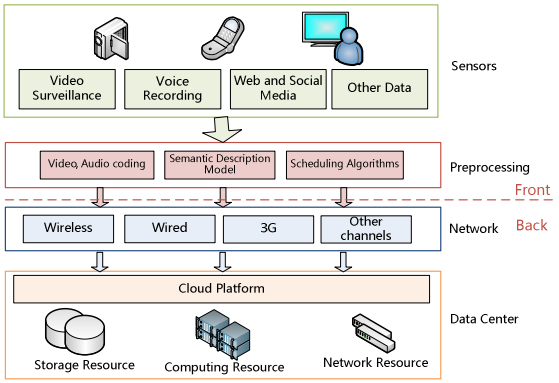

Figure 1: The “Front + Back” cloud platform for the next generation public security system.

Take the video surveillance systems for example, video data are collected by cameras, in which ARM-based processing devices are embedded to do preprocessing such as video encoding, license-plate recognition, colour recognition under semantic description models, with the results including pictures and structured description with standard format transferred to the back center, and the original video data stored into databases deployed near cameras. The center provides more resources to support deep data analysis and applications for police.

4 STORAGE, COMPUTING AND RETRIEVAL OF PUBLIC SECURITY HETEROGENEOUS BIG DATA

Data collected from types of public security systems are huge and heterogeneous, which brings great challenges for efficient storage and organization, fast retrieval and computing of data. Cloud computing technologies such as virtualization, distributed storage and computing are applied to solve these problems.

143

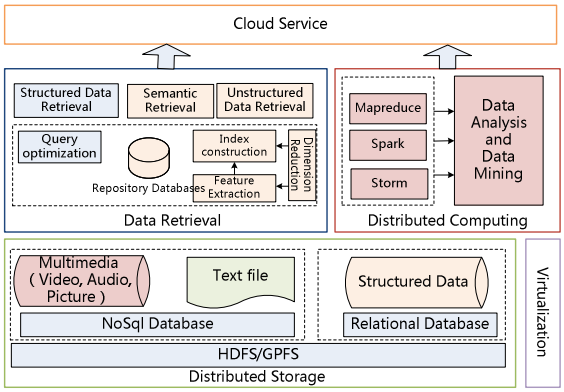

Figure2: Storage, computing and retrieval of heterogeneous big data for public security cloud platform.

4.1 Distributed Computing and Virtualization

To enhance the efficiency of tasks computing and processing, Mapreduce, Spark, Storm and other distributed processing model are applied to deal with corresponding tasks such as content analysis, semantic modelling and reasoning, other complex data analysis and mining. Take video retrieval for example, Mapreduce would be used to support the task, of which the key is represented by the time in video, and video data are divided into several parts by the key, then all tasks execute simultaneously. Virtualization is adopted to support IT resource consolidation and optimum use.

4.2 Heterogeneous Data Retrieval

In current information systems with large-scale data, data retrieval technology is basis for data access, data analysis and any other applications. Retrieval can be divided into three types: structured data retrieval, unstructured data retrieval of which pictures and text and other unstructured data are requested and semantic search, which is the results are the answers of reasoning in association with repository databases.

Strategies are applied to enhance the efficiency and availability of data retrieval. For large-scale structured data retrieval, distributed parallel database system and query optimization technology has been concerned and utilized. For retrieval of unstructured data and high-dimensional data, dimension-reduction technology is carried out during feature extraction and index construction, and the above processes and tasks could run under distributed computing frameworks.

4.3 Distributed Storage

Data collected from public security systems and sensors include structured data, video, audio, images, web pages and other unstructured data, also the output results from data processing and computing are heterogeneous. It is very critical and necessary to make useful strategizes to store and organize those data to support data analysis, data retrieval and other computing tasks effectively in different scenes.

For the structured data, parallel databases are adopted and the storage strategy is optimized according to the access frequency and characteristics of kinds of tables to satisfy the data request from kinds of applications such as statistics, retrieval, analysis and visualization. Also, for the analyzed applications, data could be stored in the Relational and NoSql databases such as Hbase by columns to support the tasks running efficiency.

For the retrieval of images and other unstructured data, Hbase model could be applied to store those data into distributed file systems such as HDFS and GPFS, so that other computing models are able to access data rapidly and efficiently.

5 PREVIOUS WORK

During 2008-2009, the third research institute of Ministry of Public Security introduced video structured description technology for the demand in video surveillance applications, and undertakes a series national science and technology major projects including the Ministry of National Science and Technology Support project, 863 smart city project and the Core Electronic Devices, High-end Generic Chips and Basic Software project. Numbers of public security intelligent video surveillance systems are carried out successfully, including VSD based road surveillance video retrieval system in Shuangliu in Chengdu, Taicang in Jiangsu Province and so on.

6 CONCLUSIONS

In this paper, we conclude a novel architecture for next generation public security system, and the “front + back” pattern is adopted to address the problems brought by the redundant construction of current public security information systems which realizes the resource consolidation of multiple IT resources, and provides unified computing and

144

Building the Cloud Platform for the Next Generation Public Security Application

storage environment for more complex data analysis and applications such as data mining and semantic reasoning. Under the architecture, we introduce cloud computing technologies such as distributed storage and computing, data retrieval of huge and heterogeneous data, provide multiple optimized strategies to enhance the utilization of resources and efficiency of tasks.

However, some other problems still exist: in what way the services could be provided to users, and it still cannot satisfy the routine detection and application for police. For example, combining crime prediction results with visualization methods is necessary for users during detection. These unsolved problems particularly merit our further study.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, in part by the National Science Foundation of China under Grant 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009.

REFERENCES

Liu,Y., Zhu,Y., Ni, Lionel M., and Xue, G. 2011. A Reliability-Oriented Transmission Service in Wireless Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, 22(12): 2100-2107.

Liu, Y., Zhang, Q., and Ni, Lionel M. 2010. Opportunity-Based Topology Control in Wireless Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, 21(3): 405-416.

Hu, C., Xu, Z. et al. 2014. Semantic Link Network based Model for Organizing Multimedia Big Data. *IEEE Transactions on Emerging Topics in Computing*, 2(3):376-387.

Xu, Z. et al. 2015. Knowle: a Semantic Link Network based System for Organizing Large Scale Online News Events. *Future Generation Computer Systems*, 43-44:40-50.

Xu, Z., Luo, X., Zhang, S., Wei, X., Mei, L., and Hu, C. 2014. Mining Temporal Explicit and Implicit Semantic Relations between Entities using Web Search Engines. *Future Generation Computer Systems*, 37:468-477.

www.nsa.gov

www.nsa.gov1.info/dni/boundless-informant.html

www.accumulo.apache.org

www.prnasia.com

145