The Intelligent Big Data Analytics Framework for Surveillance Video System

Zheng Xu, Yang Liu, Zhenyu Li and Lin Mei

Abstract Currently, with the explosion of multimedia data (image, video and audio) from remote sensors, mobile image captures, social sharing, the web, TV shows and movies, huge volume of images are being generated and consumed daily. The availability of massive images has created fundamental challenges to image processing and analysis. Big Data is a term used to refer to massive and complex datasets made up of a variety of data structures, including structured, semi-structured, and unstructured data. To address these challenges, we propose a model design methodology using collective intelligence for big data analytics. The data and data-transfer contracts then become the primary organizing constructs. With controlled data relations and timing, the system can then be built from independent agents with loosely coupled behaviors. This data-driven design technique is naturally supported by the Data Distribution Service (DDS) specification, which is a standard from the Object Management Group.

Keywords Big data · Surveillance video system · Hadoop

1 Introduction

Currently, with the explosion of multimedia data (image, video and audio) from remote sensors, mobile image captures, social sharing, the web, TV shows and movies, huge volume of images are being generated and consumed daily. The availability of massive images has created fundamental challenges to image processing and analysis. Big Data is a term used to refer to massive and complex

Z. Xu · Z. Li (✉) · L. Mei
The Third Research Institute of Ministry of Public Security, Shanghai, China

Z. Xu
Tsinghua University, Beijing, China

Y. Liu
Shanghai University, Shanghai, China

datasets made up of a variety of data structures, including structured, semi-structured, and unstructured data. Today, businesses are aware that big data can be used to generate new opportunities and process improvements through their processing and analysis. The emergence of big data has brought about a paradigm shift to many fields of computing. We have seen remarkable advances in computing power and storage capacity of big data management. But most big data systems currently in use handle data types of text or numbers. Novel and scalable data management and analytical frameworks are needed to meet the challenges posed by the big images. With the development of meteorological instrumentation and the network of surface weather stations, much more data was being collected, which leads to an extreme increase in the data's capacity. At the same time, a higher demand for easy access to all the data and new storage requirements are collected by end users [1, 2], especially for the grid data, the unstructured data. However, traditional data storage and management system (e.g., native file systems and SDBMS) [3, 4] cannot support massive data storage and processing. Cloud computing and distributed file system can give us a new resolution to these problems. Derived from Google's MapReduce and Google File System (GFS) papers, Apache Hadoop [5, 6] is an open source software framework that supports data-intensive distributed applications. Through cloud computing technology including distributes storage and computing framework, the problem brought by the huge amounts of data would be solved. Cloud computing can speed up the big data processing and storage efficiency. The design of a big data analytics system differs considerably from that of a traditional database-supported decision support system (DSS). Such a system involves more entities, data and participants; therefore, the system has special requirements in terms of data management, model design and quality of service (QoS). To address these challenges, we propose a model design methodology using collective intelligence for big data analytics. According to Wikipedia, collective intelligence (CI) is the shared or group intelligence that emerges from the collaboration, collective efforts, and competition of many individuals and appears in consensus decision making. CI systems, including multi-agent systems, complex adaptive systems, swarm intelligence and self-organizing systems, are complex by nature. The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the basic framework. Section 4 proposes a surveillance video system. Finally, we draw the conclusions of this research.

2 Related Work

Big data analytics address large volumes and distributed aggregations of various types of data. The data may be from audio, video, social networks, or web forums. Big data no longer relies on databases or data warehouses. No SQL methods, such as data management and processes in memory, are incorporated into the system. Therefore, integrating different data management mechanisms is a considerable challenge. Big data analytics models are not typically predefined due to the

presence of dynamic environments. Such models typically require iterative solutions for testing and improvement. Moreover, business processes in big data analytics systems should be flexible [7]. Participants in such models include software systems, mobile devices, web services, and humans. Building dynamic business processes that allow for cooperation amongst various participants is another challenge. QoS is also vital in big data analytics systems [8]. Jacobs [9] states that "It is easier to get the data in than out". Systems occasionally need to react to an event, such as a service outage or a change in a patient's medical condition, in real time. Another problem is that data are often incomplete; therefore, they are inferred probabilistically, and the analysis results are fuzzy. Thus, obtaining an overview of QoS properties of the system during design is the third challenge work, improve the level of intelligent video surveillance system.

3   The Intelligent Big Data Analytics Framework

From different sources, meteorological data may include observation data, forecast and service product data, mete data. Some are collected from local observation network, some are received by CMA cast and others are shared by around provinces. Data of the same category, due to different sources, may be stored in different formats. Therefore, a wide variety of diverse formats, different forms of expression, a huge amount of data, complex category, are the characteristics of meteorological data. According to Miller and Mork, big data analytics means a value chain from data to decisions through a series of processes, including data discovery, data integration, and data exploitation. These data processes are not new to software systems. However, how to implement them to support the aforementioned features in a big data analytics environment is a new challenge. In this research, we choose the multi-agent paradigm in CI. The key to the design is to separate data from behavior. Each behavior is addressed by a group of agents. The data and data-transfer contracts then become the primary organizing constructs. With controlled data relations and timing, the system can then be built from independent agents with loosely coupled behaviors. This data-driven design technique is naturally supported by the Data Distribution Service (DDS) specification, which is a standard from the Object Management Group. Triangles represent intelligent agents. The solid triangles are the administrators in a group of agents. Administrator agents attempt to create and manage other agents involved in a certain task. A group of agents that share the same goal base is called an agent platform.

   The Data Management Layer provides the basic process functions for various types of data. Different agent platforms perform different actions on SQL-like, non-SQL-like or memory-based data for storage, access and integration. Because big data systems typically require real-time functionality, such as in stream computing technology, we design several administrator agents to perform such no-storage data processes combined with traditional database operations in

distributed agent platforms. The processes in stream computing would be rapidly organized by the multi-agent mechanism as soon as the data are provided.

The Data Analysis Layer attempts to analyze data for data exploitation and decision making. The main actions on the data include query, model, analyze and visualize. Each action can be supported by a group of agents with their goal and knowledge base. These agents represent existing software systems, web services, cloud applications or any other participants in an organization. A composition of the four types of agents can be used in various business applications.

The QoS Layer provides information exchange contracts between agents. Traditional messaging designs focus on functional or operational interfaces. However, in the multi-agent system, the interface specifies the common, logically shared data model that are produced and used by an agent along with the QoS requirements as its goal, including timing, reliability, workload, and security. With such explicit QoS terms, responses to impedance mismatches can be automated, monitored, and governed.

4 Distributed Surveillance Video System Based on Hadoop

Huge small raw source files (several M or less than 1 M) with high update frequency(less than 6 min, some in seconds) should often be searched or queried in business systems, at the same time the results back to the ends users are requested to be finished in several seconds. The current data storage and download system don't find good solutions to it. In this paper, we try to first combine these small files to one big file and put it into HDFS, then begin some experiments.

Observation data. Observation data includes ground and upper-air sounding, agricultural data, radar, satellite, automatic weather station data, wind profile, GPS, microwave radiometer. Basically, Observation data is in text form, in the form of an Access database, in binary form or image format. The source data files in those categories such as automatic weather stations, microwave radiometer, road stations, atmosphere observation, radiation, in one month period, are combined into one file by using Sequence method. Objective field grid data Forecast and service product data includes four categories: numerical analysis and instructor products received from CMA cast and stored in binary format with GRIBS code, local numerical model instructor products in net cdf format, analysis products from MICAPS in micaps text format, and meteorological service product in text format (Doc, pdf, xml). Numerical model forecast products applied in the weather forecast business are from a wide variety system sources, including CMS, BMB, URUION, JAPAN and others. Therefore, there are many differences among file format, coding method, space range and resolution of source data files, which make business system burdened to parse these products to integrate application. The forecast of a single meteorology element (for examples, precipitation, temperature, wind, air

pressure, relative humidity, cloud amount and visibility) is extracted and recorded in net cdf format by interpreting them, as available objective field data for forecast operation.

According to the functions of an agent, the system contains administrator agents and executive agents. An administrator agent attempts to receive a task, make an executive plan, create executive agents, control the QoS, and send orders. The agent has a goal set and a group of calculation functions to make a plan. An executive agent attempts to realize a specific task according to orders from an administrator agent. Communication between an administrator agent and executive agent: An administrator agent sends various orders, such as creation, execution, and delete, to executive agents. This agent also sends a dataset for processing if necessary. The executive agent sends various statuses, such as finished, failed, and interrupted, to the administrator agent. This agent also sends the processed data back to the administrator agent in the data part. Communication between executive agents: If a task requires the collaboration of agents, executive agents can communicate with each other through statuses and data. Communication between administrator agents: The interactions between administrator agents aim at passing data and tasks between them.

5 Conclusions

Currently, with the explosion of multimedia data (image, video and audio) from remote sensors, mobile image captures, social sharing, the web, TV shows and movies, huge volume of images are being generated and consumed daily. The availability of massive images has created fundamental challenges to image processing and analysis. Big Data is a term used to refer to massive and complex datasets made up of a variety of data structures, including structured, semi-structured, and unstructured data. To address these challenges, we propose a model design methodology using collective intelligence for big data analytics. The data and data-transfer contracts then become the primary organizing constructs. With controlled data relations and timing, the system can then be built from independent agents with loosely coupled behaviors. This data-driven design technique is naturally supported by the Data Distribution Service (DDS) specification, which is a standard from the Object Management Group.

References

1. Luo X, Xu Z, Yu J, Chen X (2011) Building association link network for semantic link on web resources. IEEE Trans Autom Sci Eng 8(3):482–494
2. Liu Y, Ni L, Hu C (2012) A generalized probabilistic topology control for wireless sensor networks. IEEE J Sel Areas Commun 30(9):1780–1788
3. Hu C, Xu Z et al (2014) Semantic link network based model for organizing multimedia big data. IEEE Trans Emerg Top Comput 2(3):376–387
4. Liu X, Yang Y, Yuan D, Chen J (2013) Do we need to handle every temporal violation in scientific workflow systems. ACM Trans Softw Eng Methodol
5. Wang L, Tao J et al (2013) G-Hadoop: MapReduce across distributed data centers for data-intensive computing. Future Gener Comput Syst 29(3):739–750
6. Xu Z et al (2015) Knowle: a semantic link network based system for organizing large scale online news events. Future Gener Comput Syst 43–44:40–50
7. Talia D (2013) Clouds for scalable big data analytics. Computer 46(5):98–101
8. Marx V (2013) Biology: the big challenges of big data. Nature 498(7453):255–260
9. Jacobs A (2009) The pathologies of big data. Commun ACM 52(8):36–44