Mobile cloud analytics in Big data era

ZORAN BOJKOVIC, DRAGORAD MILOVANOVIC
University of Belgrade
Studentski Trg 1, 11000 Belgrade
SERBIA

Abstract: Voluminous data are generated from a variety of users and devices and are to be stored and processed in powerful data center. As such, there is a strong demand for building a network infrastructure to gather distributed and rapidly generated data and move them to data center for knowledge discovery. Big data has received considerable attention, because it can mine new knowledge for economic growth and technical innovation. Many research efforts have been directed to big data processing due to its high volume, velocity and variety, referred to as 3V. This paper first describes challenges for big data together with basic Map Reduce structure. Then it presents existing approaches for big data analytics including general architecture. The second part establishes the relation between mobile cloud and big data and provides geo-dispersed big data application, including big data in mobile cloud sensing. Some open research questions that need to be investigate conclude this work.

Key-Words: Big data analytics, mobile cloud, spatial data, wireless sensing

1 Introduction

Composed of text, image, video, audio, mobile or other forms of data collected from multiple data sets, big data is growing in size and complexity in a fast way. As a consequence, a huge volume of multi-dimension of data within a very short time period is created [1]. This is having a profound impact society and social interaction, creating at the same time opportunities for business. Big data needs vision and dialog from various disciplines such as engineers, computer scientists, statisticians, sociologists, etc. With the rapid development on the of Things (IoT), much more data is automatically generated by millions of nodes with high mobility, for example sensors carried by moving objects or vehicles. To achieve relevant information, mobile big data has to be carefully analyzed. It provides opportunities to understand behaviors and requirements of mobile users allowing the delivery of decisions for real applications.

On the other hand, network operators, will also benefit mobile big data. It should be noted that mobile big data yields many challenges to the data mining, mobile sensing and knowledge. The rapid growth of data size created a need for multi-disciplinary collaboration from industries and academics to develop new methods that can accommodate data networking, management, computational and statistical sciences to the big data analysis. A multidisciplinary approach to big data is shown in Fig. 1.

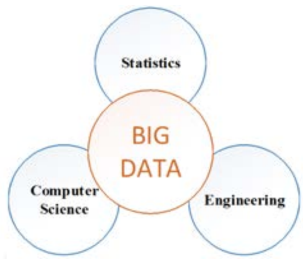

Fig. 1 Multidisciplinary relations in Big data.

This paper is organized as follows. First, we present challenges for big data followed by basic MapReduce structure. Then, we briefly review existing approach for big data analytics. The corresponding general architecture is included, too. Mobile cloud and big data together with providing geo-dispersed applications are presented. Finally, by big data in mobile cloud sensing are reported.

2 Challenges for big data

Big data is the next thing in computing. As this data cannot be processed using traditional systems, it poses numerous challenges to the researchers [2]. This raises several new challenges such as how to classify big data for multiple data sets, how to analyze big data for multiple data sets, how to analyze big data for different forms of data, how to visualize big data without loss of information. Also, privacy is one of the important concerns with big data. Many problems of big data, such as measurement, representation, compression, analysis are also faced. Big data applications introduce unprecedented challenges, while existing theories and techniques have to be extended and upgraded to serve the forthcoming real big data applications. Of course, new tools for big data applications need to be invented, for example tools for the processing of super large graph or matrix and statistical studies in low or extremely low probability events.

Technical challenges for big data include real-time analysis requirement, new storage models, as well as parallel/distributed operators for data with new *n*-dimensional array-based data structures [2]. These data need to be server-managed or cloud-managed, compared and visualized with joint efforts from hardware/software engineers, computer researchers and statisticians. Another challenge and an urgent need belongs to big data visualization. In order to lower the cost of visualization generation, new data base technologies have to be taken into account together with emerging Web-based technologies.

2.1 Basic MapReduce structure

Proposed by Google, MapReduce is well known tool for big data processing. The architecture is shown in Fig. 2. The structure is composed of two phases: map phase and reduce phase.

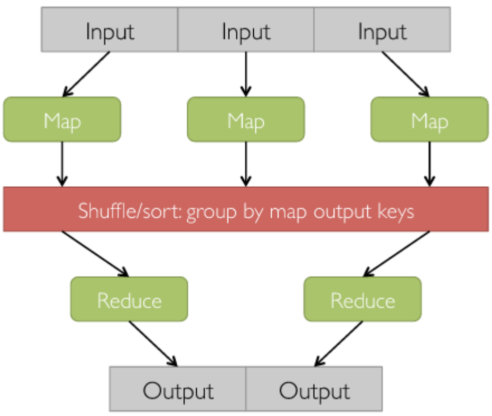

Fig. 2 Basic MapReduce processing structure.

Therefore, there exists mapper program function and reduce program function. It can be seen that in the first phase, the input data is split into blocks. The function of the mappers is to scan the data block in order to produce intermediate data. Secondly, the reduce phase starts from a shuffle sub-phase. It shuffles intermediate data, run by reducers. They are moved to the corresponding nodes, for reducing. As for the reducers, they process the shuffled data generating the final results. When dealing with more complicated problems, multiple map and reduce processes can be used.

2.2 MapReduce implementations

There exists a number of practical MapReduce implementations. Nowadays the open source Hadoop is the most widely used. Generally, Hadoop and other MapReduce-like tools run on dedicated server cluster. Therefore, studies have mainly focused on optimization using such infrastructures, enabling the analysis of massive data. In order to be in position to realize big data analytics, a set of requirements have to be taken into consideration. Here, expected results, budget, response time, reliability, accuracy, are included [3]. To enable big data analytics, necessary functions are grouped into three layers: infrastructure layer, platform layer, and software layer.

In the infrastructure layers a network of machines will be set up for computing, storage, and communications. A trend for big data analytics is to use the cloud as the infrastructure because of scalability and economics. A cloud consists of usable and accessible virtualized resources. From the point of view of cloud usage, a cloud is a public cloud if the cloud service provider offers services to the general public.

In the platform layer, Hadoop is the computing paradigm. Currently, is is used in many major big data analyzers including Internet companies *Yahoo*, *Amazon*, *Facebook*, *Twitter* and many others data analytics companies [3]. As a Java implementation of MapReduce, Hadoop supports large-scale data analytics applications. However, MapReduce does not fit iterative algorithms that need to handle the dependency between data over iterations. Spark [4], as in-memory computing framework is designed to overcome Hadoop's shortages in iterative operations. It introduces a real-only data structure-resilient distributed data sets. During the whole iterative process, intermediate results can be cached in memory. GraphLab has evolved to introduce a new approach to distributed graph placement. It should be noted that each platform may target different applications.

The big data analyzer decides which of algorithms to use in the first step of data analysis. However, even for a given algorithm, there exist different implementation methods with different languages and on different analytics platform. Thus for a big data analyzer in order to have reliable and accurate data, it is necessary for a big data analyzer to understand the details of algorithms.

2.3 General architecture of analytics

Block scheme of general architecture consists of multi-source big data collecting, distributed big data storing and finally intra/inter big data processing. Big data collection is characterized by high volume, high velocity and high variety (3V). General architecture of big data analytics is shown in Fig. 3.

Fig. 3 General architecture of Big data analytics.

The challenges are:

- how to efficiently store and organize high-volume data,
- how to quickly process streaming and near-real-time data, and
- how to accurately analyze structured and unstructured data in order to maximize the value of big data.

Since big data collecting is pervasive, it is not efficient to move high volumes of collected data around for centralized storing. Thus, distributed big data is proposed [5]. It means that big data will be stored and organized in their original location. Big data should be processed in parallel such that new knowledge and innovation can be mined in a reasonable amount of time. Big data processing is divided in two groups: intra processing, and inter processing.

In inter big data processing, all data belong to the same organization. On the other side, if big data are part of different organizations, it will be inter big data processing. It should be emphasized that inter big data processing will be more challenging because big data sharing will first be executed before processing. Also, many new security and privacy issues will arise during the duration of big data sharing.

3 Mobile cloud and Big data

3.1 Spatial Big data application

Mobile services (*Google Maps*, Navigation service) are big data applications. Because of the big data set and the data update rate is fast, benefits and convenience are provided [6]. Thus mobile cloud is used for supporting geo-dispersed big data applications. Framework based on MapReduce for simple and complex operations is proposed to provide better support and for geo-dispersed big data [7]. Block scheme of mobile cloud architecture is shown in Fig. 4. It consists of several cloudlets and a central cloud. Advanced MapReduce framework (AMF) employs cooperative processing in the mobile cloud and MapReduce to process geo-dispersed big data.

AMF extracts the required multiple inputs from geo-dispersed big data in parallel. After that, required extracted multiple inputs from different cloud nodes are aggregated and processed after performing complex operations. Then, creating the final results, they are sent to the user. Different data aggregation schemes are used for different application requirements.

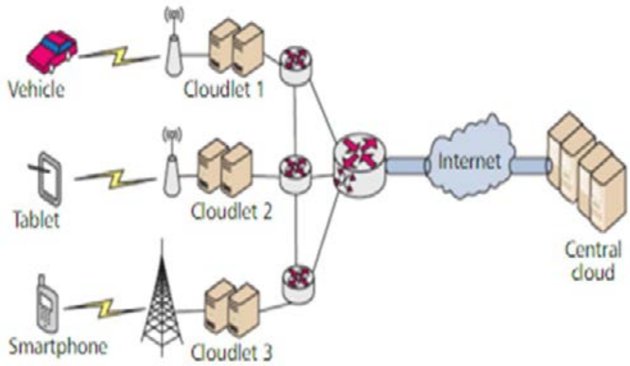

Fig. 4 Mobile cloud architecture for providing geo-dispersed Big data applications.

As for the mobile cloud architecture, it consists of several cloudlets and a central cloud which stores part of the big data, while the cloudlets have large amounts of thresh data as a part of big data which are uploaded to the central cloud periodically to update the data set. The central cloud has got sufficient computational resources to process all of the big data. Note, that we have large amounts of fresh data from cloudlets to the central clod, while communication delays to mobile users is short. Thus, a cloudlet should be utilized to assist in performing operations on geo-dispersed big data and reduce response time.

3.2 Mobile wireless Big data application

Mobile wireless big data applications together with advanced technologies in mobile networking, are emerging today. For example, *Nike+* provides important services for users with wireless connected sensing devices and smartphones [8]. Company *Nike* enhances its products, such as shoes, with built in sensors that continuously track the user's movement in the period their workouts. The collected data provide users with their instant information such as their pace, GPS position, distance moved. Of course, *Nike*+ apps installed in the users' smartphone collect data from the sensors using wireless connections [9]. *Nike*+ has become a big data platform that collects, stores and processes data generated from a huge number of users.

3.3 Big data in mobile cloud sensing

Mobile cloud sensing combines mobile sensing, cloud computing and big data to obtain process and predict mobile sensor data. One of the bottleneck for data centric sensing such as video surveillance, image sharing, etc is network bandwidth. As more sensors are integrated into mobile devices, the generated data by them cloud exceed the network capacity. A network infrastructure has to be applied to support the data capacity as well as connection request. Mobile cloud sensing brings together cloud computing services and mobile sensing features. Also, mobile cloud sensing contributes mobile devices access to resources of cloud computing. On the other hand, cloud computing infrastructure is enabled to obtain real world data from mobile sensing devices. The mobile cloud sensing architecture with the building components is shown in Fig. 5 [9].

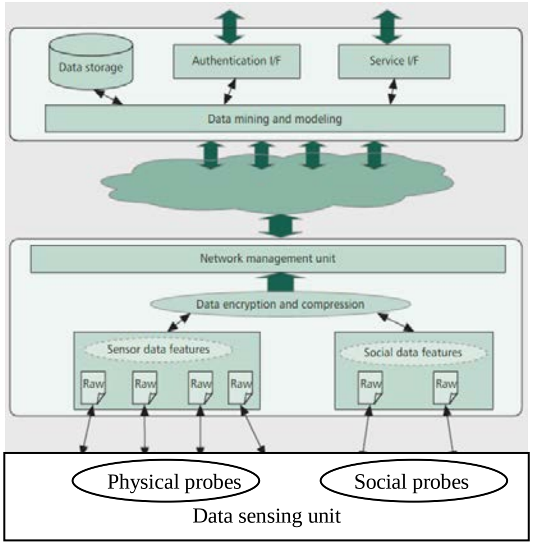

Fig. 5 Essential building blocks of a mobile cloud sensing system.

The data sensing unit consists of physical sensing probes and social sensing probes. Physical sensing probes include smartphones, tablets, and wearable devices (smartwatch, smart glasses, smartbracelets), while social sensing probes are posts on social networks (*Facebook*, *Twitter*). Physical sensing unit are raw format such as accelerometer data, ambient light strength, pulse rate, digital image, audio data.

Other part of sensing data is from the social sensing probes. Data preprocessing unit examines the raw data from sensors and social networks, extracting corresponding features encrypted and compressed to minimize the data bandwidth and protect the data. Network management unit can be optimized to make sensing data throughput larger as well as to make the integration of 5G network interfaces easier and faster. Cloud platforms have sufficient storage for sensing data. Data from sensing sources converge here, while features are fed to the specific machine learning tasks to be interpreted. Results are stored on the cloud for accessing. Data authentication and service interfaces interact directly with end users.

4 Conclusion

Big data presents a new era of information generation and processing. However, research work of big data processing in the mobile cloud remains in its infancy in spite of the fact that cloud computing is a popular infrastructure that has the resources for big data processing. On the other hand mobile cloud computing is becoming important part of big data applications. In connection with this mobile opportunistic networks can function as a complementary alternative infrastructures for supporting the emerging big data applications. As a technology that makes an intelligent and smart world possible through mobile devices mobile cloud sensing is changing the way we live. New data systems and technologies are required to handle mobile big data in a highly scalable, cost effective and fault tolerant fashion.

References:

[1] J.Zhang and M.L.Huang, Data behaviors model for Big data signal analytics, *International Journal of Big data intelligence*, vol. 3, no.1, pp. 1-17, 2016.

[2] H.Fang, et al., A survey of Big data research, *IEEE Network*, vol.29, no.5, pp.6-9, Sept/Oct. 2015.

[3] H.Li and Sh.Meng, Big provision: a provision framework Big data analytics, *IEEE Network*, vol.29, no.5, pp.50-56, Sept/Oct. 2015.

[4] *Spark engine for large-scale data processing* (available online http://spark-project.org)

[5] R.Lu, et al., Toward efficient and privacy-preserving computing in Big data era, *IEEE Network*, vol.28, no.4, pp.46-59, July/Aug. 2014.

[6] S.Shekhar, et al., Spatial Big data challenges intersecting mobility and cloud computing, in *Proc. ACM International Workshop Data engineering for Wireless and mobile access*, Scottsdale, AZ, pp.1-6, 2012.

[7] H.Zhang, et al., Processing Geo-dispersed Big data in an advanced Map Reduce framework, *IEEE Network*, vol.29, no.5, pp.24-30, Sept/Oct. 2015.

[8] X.Yi, et al., Building a network highway for Big data: architecture and challenges, *IEEE Network*, vol.28, no.4, pp.5-13, July/Aug. 2014.

[9] G.Han, Sh.Liang, H.Zhang, Mobile cloud sensing, Big data and 5G networks make an intelligent and smart world, *IEEE Network*, vol.29, no.2, pp.40-45, March/April 2015.