ISSN:0975-9646

M.H.Padgavankar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2218-2223

Big Data Storage and Challenges

M.H.Padgavankar[1], Dr.S.R.Gupta[2]

[1]*ME, CSE, PRMIT&R, Amravati, Maharashtra, India*
[2]*Assistant Professor, CSE, PRMIT&R, Amravati, Maharashtra, India*

Abstract — Now days the big data has became the most difficult problem in the Industrial ,Science ,Education sector. Here we discussing the storage problems in these sectors. how the data is explode in the recent years. And how we are dealing with the massive amount of data in our sectors. We are also focusing on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, we introduce the general background, discuss the technical challenges, and Opportunities in the big data. we have seen the massive size of the data in industries and in the public and social sectors. These discussions aim to provide a comprehensive overview and big-picture to readers of this exciting area. This survey is concluded with a discussion of open problems and future directions .

Keywords - Big data, Internet of things , Data center ,Hadoop , Big data analysis.

1.INTRODUCTION

A wide range of project businesses are involved in the life-cycle of engineer-to-order goods such as buildings. We are awash in a flood of data today. In a broad range of application areas, data is being collected at extraordinary scale. Decisions that previously were based on guesswork, or on painstakingly constructed models of reality, can now be made based on the data itself. Such Big Data analysis now drives nearly every aspect of our modern society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences.

Scientific research has been revolutionized by Big Data. The field of Astronomy is being transformed from one where taking pictures of the sky was a large part of an astronomer's job to one where the pictures are all in a database already and the astronomer's task is to find interesting objects and phenomena in the database. Imagine a world in which we have access to a huge database where we collect every detailed measure of every student's academic performance. This data could be used to design the most effective approaches to education, starting from reading, writing, and math, to advanced, college-level, courses. We are far from having access to such data, but there are powerful trends in this direction. In particular, there is a strong trend for massive Web deployment of educational activities, and this will generate an increasingly large amount of detailed data about students' performance. Over the past 20 years, data has increased in a large scale in various fields. According to a report from International Data Corporation (IDC), in 2011, the overall created and copied data volume in the world was 1.8ZB (≈ 1021*B*)[1].which increased by nearly nine times within five years. This figure will double at least every other two years in the near future. Recently, industries become interested in the high potential of big data, and many government agencies announced major plans to accelerate big data research and applications [2]. In addition, issues on big data are often covered in public media, such as *The Economist* [3, 4], *New York Times* [5], and *National Public Radio* [6, 7]. Two premier scientific journals, *Nature* and *Science*, also opened special columns to discuss the challenges and impacts of big data [8, 9]. The era of big data has come beyond all doubt [10]. The capacities of the IT architectures and infrastructure of existing enterprises, and its realtime requirement will also greatly stress the available computing capacity.

The increasingly growing data cause a problem of how to store and manage such huge heterogeneous datasets with moderate requirements on hardware and software infrastructure. Nowadays, big data related to the service of Internet companies grow rapidly. For example, Google processes data of hundreds of Petabyte (PB), Facebook generates log data of over 10 PB per month, Baidu, a Chinese company, processes data of tens of PB, and Taobao, a subsidiary of Alibaba generates data of tens of Terabyte (TB) for online trading per day.

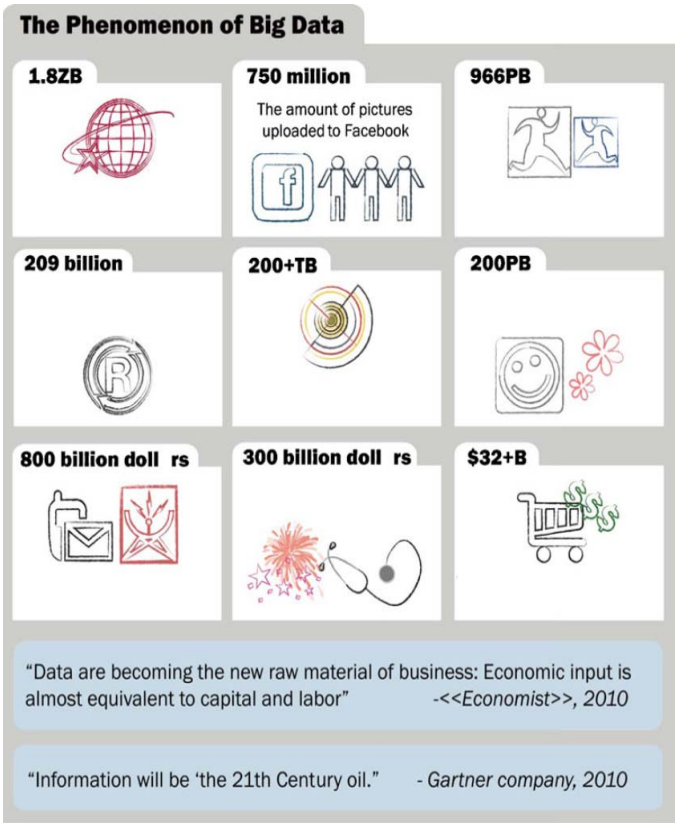

Figure 1 Increasing big data

www.manaraa.com

2.DEFINING THE BIG DATA

Big data is an abstract concept. Apart from masses of data, it also has some other features, which determine the difference between itself and "massive data" or "very big data."

OR

"Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization."

Big Data refers to new database management and analytical approaches developed for analyzing, storing, and manipulating large or complex data. Investments in Big Data include those in human resources (e.g., data scientists) and in business and technology solutions, including database management platforms (e.g., Hadoop, IBM/Netezza), analytics and visualization capabilities (e.g., Revolution R) or text-processing and real-time streaming solutions.

Big Data refers to datasets whose size are beyond the ability of typical database software tools to capture, store, manage and analyse. There is no explicit definition of how big a dataset should be in order to be considered Big Data. New technology has to be in place to manage this Big Data phenomenon. IDC defines Big Data technologies as a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high velocity capture, discovery and analysis.Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or does not fit the structures of existing database architectures. To gain value from these data, there must be an alternative way to process it.

In the late 1970s, the concept of "database machine" emerged, which is a technology specially used for storing and analyzing data. With the increase of data volume,the storage and processing capacity of a single mainframe computer system became inadequate. In the 1980s, people proposed "share nothing," a parallel database system, to meet the demand of the increasing data volume.[11] The share nothing system architecture is based on the use of cluster and every machine has its own processor, storage, and disk. Tera data system was the first successful commercial parallel database system. Such database became very popular lately. On June 2, 1986, a milestone event occurred when Teradata delivered the first parallel database system with the storage capacity of 1TB to Kmart to help the large-scale retail company in North America to expand its data warehouse.[12] In the late 1990s, the advantages of parallel database was widely recognized in the database field. However, many challenges on big data arose. With the development of Internet servies, indexes and queried contents were rapidly growing. Therefore, search engine companies had to face the challenges of handling such big data. Google created GFS[13] and MapReduce [14]programming models to cope with the challenges brought about by data management and analysis at the Internet scale. In addition, contents generated by users, sensors, and other ubiquitous data sources also feuled the overwhelming dataflows, which required a fundamental change on the computing architecture and large-scale data processing mechanism.

At present, data has become an important production factor that could be comparable to material assets and human capital. As multimedia, social media, and IoT are developing, enterprises will collect more information, leading upgrading huge volume and heterogeneity of big data. The research community has proposed some solutions from different perspectives. For example, cloud computing is utilized to meet the requirements on infrastructure for big data, e.g., cost efficiency, elasticity, and smooth downgrading.

3. CHARACTERISTICS OF BIG DATA

The characteristics of the big data depends on the three factors which includes Data Velocity, Data Volumn and Data Variety**.** Big Data is not just about the size of data but also includes data variety and data velocity. these are the three V's of the Big data.[1]

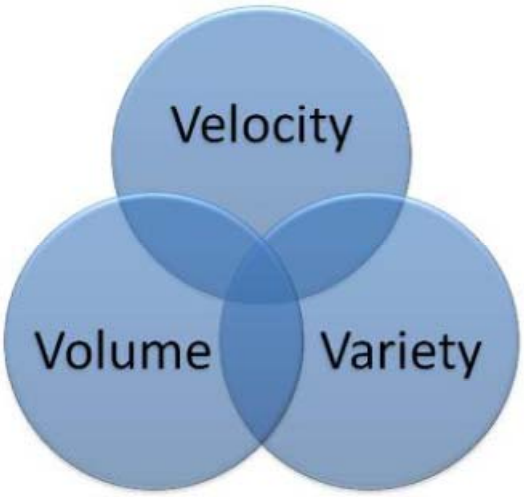

Figure 2 The V's of the Big Data

Variety- The next aspect of Big Data is its variety. This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts. This helps the people, who are closely analyzing the data and are associated with it, to effectively use the data to their advantage and thus upholding the importance of the Big Data.

Velocity- The term 'velocity' in this context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development.

Variability- This is a factor which can be a problem for those who are analyse the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Complexity- Data management can become a very complex process, especially when large volumes of data come from multiple sources. These data need to be linked, connected

www.manaraa.com

and correlated in order to be able to grasp the information that is supposed to be conveyed by these data. This situation, is therefore, termed as the 'complexity' of Big Data.

Data comes mainly in two forms-

1. Structured, and
2. Unstructured Data (there are also semi-structured data – eg. XML)

Structured data has semantic meaning attached to it whereas Unstructured data has no latent meaning.

The growth in data that we are referring is most unstructured data. Below are few examples of unstructured data –

1. Calls, text, tweet, net surf, browse through various websites each day and exchange messages via several means.
2. Social media usage my several million people for exchanging data in various forms also forms a part of Big Data.
3. Transactions made through card for various payment issues in large numbers every second across the world also constitutes the Big Data.

Hope this posts gave you enough of infomation about Big Data and in future posts, we will be looking at – Applications of Big Data i.e. Big Data Analytics, Careers in Big Data – From Software Engineer to becoming a Data Scientist, Hadoop and Applications.

Handling the three Vs helps organisations extract the value of Big Data. The value comes in turning the three Vs into the three Is:

1. Informed intuition: predicting likely future occurrences and what course of actions is more likely to be successful.
2. Intelligence: looking at what is happening now in real time (or close to real time) and determining the action to take.
3. Insight: reviewing what has happened and determining the action to take.

4. CHALLENGES OF BIG DATA

The sharply increasing data deluge in the big data era brings about huge challenges on data acquisition, storage, management and analysis. Traditional data managementand analysis systems are based on the relational database management system (RDBMS). However, such RDBMSs only apply to structured data, other than semi-structured or unstructured data. In addition, RDBMSs are increasingly utilizing more and more expensive hardware. It is apparently that the traditional RDBMSs could not handle the huge volume and heterogeneity of big data. The research community has proposed some solutions from different perspectives.

For example, cloud computing is utilized to meet the requirements on infrastructure for big data, e.g., cost efficiency, elasticity, and smooth upgrading/downgrading. For solutions of permanent storage and management of large-scale disordered datasets, distributed file systems and NoSQL databases are good choices. Such programming frameworks have achieved great success in processing clustered tasks, especially for webpage ranking. Various big data applications can be developed based on these innovative technologies or platforms. Moreover, it is non-trivial to deploy the big data analysis systems.

Some literature[15-16] discuss obstacles in the development of big data applications. The key challenges are listed as follows:

– *Data representation*: many datasets have certain levels of heterogeneity in type, structure, semantics, organization, granularity, and accessibility. Data representation aims to make data more meaningful for computer analysis and user interpretation. Nevertheless, an improper data representation will reduce the value of the original data and may even obstruct effective data analysis. Efficient data representation shall reflect data structure, class, and type, as well as integrated technologies, so as to enable efficient operations on different datasets.

– *Redundancy reduction and data compression*: generally, there is a high level of redundancy in datasets. Redundancy reduction and data compression is effective to reduce the indirect cost of the entire system on the premise that the potential values of the data are not affected. For example, most data generated by sensor networks are highly redundant, which may be filtered and compressed at orders of magnitude.

– *Data life cycle management*: compared with the relatively slow advances of storage systems, pervasive sensing and computing are generating data at unprecedented rates and scales. We are confronted with a lot of pressing challenges, one of which is that the current storage system could not support such massive data. Generally speaking, values hidden in big data depend on data freshness. Therefore, a data importance principle related to the analytical value should be developed to decide which data shall be stored and which data shall be discarded.

– *Analytical mechanism*: the analytical system of big data shall process masses of heterogeneous data within a limited time. However, traditional RDBMSs are strictly designed with a lack of scalability and expandability, which could not meet the performance requirements. Non-relational databases have shown their unique advantages in the processing of unstructured data and started to become mainstream in big data analysis. Even so, there are still some problems of non-relational databases in their performance and particular applications.

We shall find a compromising solution between RDBMSs and non-relational databases. For example, some enterprises have utilized a mixed database architecture that integrates the advantages of both types of database (e.g., Facebook and Taobao). More research is needed on the in-memory database and sample data based on approximate analysis.

– *Data confidentiality*: most big data service providers or owners at present could not effectively maintain and analyze such huge datasets because of their limited capacity. They must rely on professionals or tools to analyze such data, which increase the potential safety risks. For example, the transactional dataset generally includes a set of complete operating data to drive key business processes. Such data contains details of the lowest

www.manaraa.com

granularity and some sensitive information such as credit card numbers. Therefore, analysis of big data may be delivered to a third party for processing only when proper preventive measures are taken to protect such sensitive data, to ensure its safety.

– *Energy management*: the energy consumption of mainframe computing systems has drawn much attention from both economy and environment perspectives.With the increase of data volume and analytical demands, the processing, storage, and transmission of big data will inevitably consume more and more electric energy. Therefore, system-level power consumption control and management mechanism shall be established for big data while the expandability and accessibility are ensured.

– *Expendability and scalability*: the analytical system of big data must support present and future datasets. The analytical algorithm must be able to process increasingly expanding and more complex datasets.

– *Cooperation*: analysis of big data is an interdisciplinary research, which requires experts in different fields cooperate to harvest the potential of big data.

A comprehensive big data network architecture must be established to help scientists and engineers in various fields access different kinds of data and fully utilize their expertise, so as to cooperate to complete the analytical objectives.

5. BIG DATA GENERATION AND ACQUISITION

Can be generally divided into four phases: data generation, data acquisition, data storage, and data analysis. If we take data as a raw material, data generation and data acquisition are an exploitation process, data storage is a storage process, and data analysis is a production process that utilizes the raw material to create new value.[9]

5.1. Data Generation:

Data generation is the first step of big data. Given Internet data as an example, huge amount of data in terms of searching entries, Internet forum posts, chatting records, and microblog messages, are generated. Those data are closely related to people's daily life, and have similar features of high value and low density. Such Internet data may be valueless individually, but, through the exploitation of accumulated big data, useful information such as habits and hobbies of users can be identified, and it is even possible to forecast users' behaviors and emotional moods. Moreover, generated through longitudinal and/or distributed data sources, datasets are more large-scale, highly diverse, and complex. Such data sources include sensors, videos, click streams, and/or all other available data sources.

At present, main sources of big data are the operation and trading information in enterprises, logistic and sensing information in the IoT, human interaction information and position information in the Internet world, and data generated in scientific research, etc. The information far surpasses the capacities of IT architectures and infrastructures of existing enterprises, while its real time requirement also greatly stresses the existing computing capacity.

5.2. Big Data Acquisition:

As the second phase of the big data system, big data acquisition includes data collection, data transmission, and data pre-processing. During big data acquisition, once we collect the raw data, we shall utilize an efficient transmission mechanism to send it to a proper storage management system to support different analytical applications. The collected datasets may sometimes include much redundant or useless data, which unnecessarily increases storage space and affects the subsequent data analysis. For example, high redundancy is very common among datasets collected by sensors for environment monitoring. Data compression technology can be applied to reduce the redundancy. Therefore, data pre-processing operations are indispensable to ensure efficient data storage and exploitation.

5.3. Big data storage:

The explosive growth of data has more strict requirements on storage and management. In this section, we focus on the storage of big data. Big data storage refers to the storage and management of large-scale datasets while achieving reliability and availability of data accessing. We will review important issues including massive storage systems, distributed storage systems, and big data storage mechanisms. On one hand, the storage infrastructure needs to provide information storage service with reliable storage space; on the other hand, it must provide a powerful access interface for query and analysis of a large amount of data.

Traditionally, as auxiliary equipment of server, data storage device is used to store, manage, look up, and analyze data with structured RDBMSs. With the sharp growth of data, data storage device is becoming increasingly more important, and many Internet companies pursue big capacity of storage to be competitive. Therefore, there is a compelling need for research on data storage.

Storage system for massive data.Various storage systems emerge to meet the demands of massive data. Existing massive storage technologies can be classified as Direct Attached Storage (DAS) and network storage, while network storage can be further classified into Network Attached Storage (NAS) and Storage Area Network (SAN).

In DAS, various harddisks are directly connected with servers, and data management is server-centric, such that storage devices are peripheral equipments, each of which takes a certain amount of I/O resource and is managed by an individual application software. For this reason, DAS is only suitable to interconnect servers with a small scale. However, due to its low scalability, DAS will exhibit undesirable efficiency when the storage capacity is increased, i.e., the upgradeability and expandability are greatly limited. Thus, DAS is mainly used in personal computers and small-sized servers.

Network storage is to utilize network to provide users with a union interface for data access and sharing. Network storage equipment includes special data exchange equipments, disk array, tap library, and other storage media, as well as special storage software. It is characterized with strong expandability.

NAS is actually an auxillary storage equipment of a network. It is directly connected to a network through a hub

www.manaraa.com

or switch through TCP/IP protocols. In NAS, data is transmitted in the form of files. Compared to DAS, the I/O burden at a NAS server is reduced extensively since the server accesses a storage device indirectly through a network.

While NAS is network-oriented, SAN is especially designed for data storage with a scalable and bandwidth intensive network, e.g., a high-speed network with optical fiber connections. In SAN, data storage management is relatively independent within a storage local area network, where multipath based data switching among any internal nodes is utilized to achieve a maximum degree of data sharing and data management.

From the organization of a data storage system, DAS, NAS, and SAN can all be divided into three parts: (i) disc array: it is the foundation of a storage system and the fundamental guarantee for data storage; (ii) connection and network sub-systems, which provide connection among one or more disc arrays and servers; (iii) storage management software, which handles data sharing, disaster recovery, and other storage management tasks of multiple servers.

5.4 **The Data analysis:**

The analysis of big data mainly involves analytical methods for traditional data and big data, analytical architecture for big data, and software used for mining and analysis of big data. Data analysis is the final and the most important phase in the value chain of big data, with the purpose of extracting useful values, providing suggestions or decisions.

Different levels of potential values can be generated through the analysis of datasets in different fields. However, data analysis is a broad area, which frequently changes and is extremely complex. In this section, we introduce the methods, architectures and tools for big data analysis.

5.4.1 Traditional Data Analysis**:**

Traditional data analysis means to use proper statistical methods to analyze massive data, to concentrate, extract, and refine useful data hidden in a batch of chaotic datasets, and to identify the inherent law of the subject matter, so as to maximize the value of data. Data analysis plays a huge guidance role in making development plans for a country, understanding customer demands for commerce, and predicting market trend for enterprises. Big data analysis can be deemed as the analysis technique for a special kind of data.

Therefore, many traditional data analysis methods may still be utilized for big data analysis. Several representative traditional data analysis methods are examined in the following, many of which are from statistics and computer science.

– *Cluster Analysis*: is a statistical method for grouping objects, and specifically, classifying objects according to some features. Cluster analysis is used to differentiate objects with particular features and divide them into some categories (clusters) according to these features, such that objects in the same category will have high homogeneity while different categories will have high heterogeneity. Cluster analysis is an unsupervised study method without training data.

– *Factor Analysis*: is basically targeted at describing the relation among many elements with only a few factors, i.e., grouping several closely related variables into a factor, and the few factors are then used to reveal the most information of the original data.

– *Correlation Analysis*: is an analytical method for determining the law of relations, such as correlation, correlative dependence, and mutual restriction, among observed phenomena and accordingly conducting forecast and control. Such relations may be classified into two types: (i) function, reflecting the strict dependence relationship among phenomena, which is also called a definitive dependence relationship; (ii) correlation, some undetermined or inexact dependence relations, and the numerical value of a variable may correspond to several numerical values of the other variable, and such numerical values present a regular fluctuation surrounding their mean values.

– *Regression Analysis*: is a mathematical tool for revealing correlations between one variable and several other variables. Based on a group of experiments or observed data, regression analysis identifies dependence relationships among variables hidden by randomness. Regression analysis may make complex and undetermined correlations among variables to be simple and regular. According to timeliness requirements, big data analysis can be classified into real-time analysis and off-line analysis.

– *Real-time analysis*: is mainly used in E-commerce and finance. Since data constantly changes, rapid data analysis is needed and analytical results shall be returned with a very short delay. The main existing architectures of real-time analysis include (i) parallel processing clusters using traditional relational databases, and (ii) memory-based computing platforms. For example, Greenplum from EMC and HANA from SAP are both real-time analysis architectures.

– *Offline analysis*: is usually used for applications without high requirements on response time, e.g., machine learning, statistical analysis, and recommendation algorithms. Offline analysis generally conducts analysis by importing logs into a special platform through data acquisition tools. Under the big data setting, many Internet enterprises utilize the offline analysis architecture based on Hadoop in order to reduce the cost of data format conversion and improve the efficiency of data acquisition. Examples include Facebook's open source tool Scribe, LinkedIn's open source tool Kafka, Taobao's open source tool Time tunnel, and Chukwa of Hadoop , etc. These tools can meet the demands of data acquisition and transmission with hundreds of MB per second.

Analysis at different levels Big data analysis can also be classified into memory level analysis, Business Intelligence (BI) level analysis, and massive level analysis, which are examined in the following.

– *Memory-level analysis*: is for the case where the total data volume is smaller than the maximum memory of a cluster. Nowadays, the memory of server cluster surpasses hundreds of GB while even the TB level is common. Therefore, an internal database technology may be used, and hot data shall reside in the memory so as to improve

www.manaraa.com

the analytical efficiency. Memory-level analysis is extremely suitable for real-time analysis. Mongo DB is a representative memory-level analytical Architecture. With the development of SSD (Solid-State Drive), the capacity and performance of memory-level data analysis has been further improved and widely applied.

– *BI analysis*: is for the case when the data scale surpasses the memory level but may be imported into the BI analysis environment. The currently, mainstream BI products are provided with data analysis plans to support the level over TB.

– *Massive analysis*: is for the case when the data scale has completely surpassed the capacities of BI products and traditional relational databases. At present, most massive analysis utilize HDFS of Hadoop to store data and use MapReduce for data analysis. Most massive analysis belongs to the offline analysis category.

6. New Opportunities in Big Data

Since the Internet's introduction, we've been steadily moving from text-based communications to richer data that include images, videos, and interactive maps as well as associated metadata such as geo location information and time and date stamps. Twenty years ago, ISDN lines couldn't handle much more than basic graphics, but today's high-speed communication networks enable the transmission of storage-intensive data types. [17]

For instance, smartphone users can take high-quality photographs and videos and upload them directly to social networking sites via Wi-Fi and 3G or 4G cellular networks. We've also been steadily increasing the amount of data captured in bidirectional interactions, both people-to-machine and machine-to-machine, by using telematics and telemetry devices in systems of systems. Of even greater importance are e-health networks that allow for data merging and sharing of high-resolution images in the form of patient x-rays, CT scans, and MRIs between stakeholders.

Advances in data storage and mining technologies make it possible to preserve increasing amounts of data generated directly or indirectly by users and analyze it to yield valuable new insights. For example, companies can study consumer purchasing trends to better target marketing. In addition, near-real-time data from mobile phones could provide detailed characteristics about shoppers that help reveal their complex decision-making processes as they walk through malls.

Big data can expose people's hidden behavioral patterns and even shed light on their intentions. More precisely, it can bridge the gap between what people want to do and what they actually do as well as how they interact with others and their environment. This information is useful to government agencies as well as private companies to support decision making in areas ranging from law enforcement to social services to homeland security. It's particularly of interest to applied areas of situational awareness and the anticipatory approaches required for near-real-time discovery.

7. Conclusion

Here we have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. we review the background and state-of-the-art of big data. we introduce the general background of big data and review related technologies, such as could computing, IoT, data centers, and Hadoop. Then we focus on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, we introduce the general background, discuss the technical challenges, and review the latest advances.

References

1. Gantz J, Reinsel D (2011) Extracting value from chaos. IDC iView, pp 1–12
2. Fact sheet: Big data across the federal government (2012). http:// www.whitehouse.gov/sites/default/files/microsites/ostp/big data fact sheet 3 29 2012.pdf
3. Cukier K (2010) Data, data everywhere: a special report on managing information. Economist Newspaper
4. Drowning in numbers - digital data will flood the planet- and help us understand it better (2011). http://www.economist.com/blogs/ dailychart/2011/11/bigdata-0
5. Lohr S (2012) The age of big data. New York Times, pp 11
6. Yuki N (2011) Following digital breadcrumbs to big data gold.http://www.npr.org/2011/11/29/142521910/thedigitalbreadcrumbsthat-lead-to-big-data
7. Yuki N The search for analysts to make sense of big data (2011).http://www.npr.org/2011/11/30/142893065/the-searchforanalyststo-make-sense-of-big-data8. Big data (2008). http://www.nature.com/news/specials/bigdata/index.html
9. Special online collection: dealing with big data (2011). http://www.sciencemag.org/site/special/data/
10. Manyika J, McKinsey Global Institute, Chui M, Brown B,Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data:the next frontier for innovation, competition, and productivity.McKinsey Global Institute
11. DeWitt D, Gray J (1992) Parallel database systems: the future of high performance database systems. Commun ACM35(6):85–98
12. Walter T (2009) Teradata past, present, and future. UCI ISG lecture series on scalable data management
13. Ghemawat S, Gobioff H, Leung S-T (2003) The google file system. In: ACM SIGOPS Operating Systems Review, vol 37. ACM, pp 29–43
14. Dean J, Ghemawat S (2008) Mapreduce: simplified data processing on large clusters. Commun ACM 51(1):107–113
15. Labrinidis A, Jagadish HV (2012) Challenges and opportunities with big data. Proc VLDB Endowment 5(12):2032–2033
16. Agrawal D, Bernstein P, Bertino E, Davidson S, Dayal U, FranklinM, Gehrke J, Haas L, Halevy A, Han J et al (2012) Challenges and opportunities with big data. A community white paper developed by leading researches across the United States
17. Big Data: New Opportunities and New Challenges Katina Michael university of Wollongong, Keith W. Miller university of Missouri-St.Louis

www.manaraa.com