International Journal of Technology and Engineering System (IJTES)
Vol 7. No.4 2015 Pp. 283-290
©gopalax Journals, Singapore
available at : www.ijcns.com
ISSN: 0976-1345

A SURVEY ON CLOUD SUSTAINED BIG DATA COMPUTING

C. VIJAYAKUMARAN, M. JOHARAN BEEVI
Assistant Professor, PG Student,
Department of Computer Science and Engineering,
Valliammai Engineering College, Chennai, India.
c_vijayakumaran@yahoo.com, joirfan23@gmail.com

ABSTRACT

The big data flows in World Wide Web (WWW) will lead to knowledge revolutions in all areas of Information and Communication Technologies (ICT) sectors. Big data is a word applied to large, complex and dynamic data sets for which there is a obligation to capture, manage and process the data set in its totality, such that it is not possible to efficiently process the data using traditional software tools and analytic techniques. Data-Intensive Scientific Discovery (DISD), also known as Big Data problems influence large number of fields such as economic, business activities, national security and many scientific researches. Big Data is extremely useful in businesses and scientific researches, which provide high prospects to make great progresses in many fields. In future, many competitive technologies and business productivity will fall into the Big Data explorations. This paper discusses tactics and situations for carrying out analytics on Clouds for Big Data applications. It also discusses the underlying methods and techniques to handle the Big Data explorations in related applications.

Keywords: Big Data, World Wide Web (WWW), Information and Communication Technologies (ICT), Data-Intensive Scientific Discovery (DISD) and Clouds.

I INTRODUCTION

Data Analytics increase businesses productivity, thus many organizations usually do analytical activities such as sales, marketing, risk management, fraud detection and customer support to overcome business competitions and also to improve their business strategies and decisions. However, many small organizations suffer from too expensive, complex and resource intensive Big Data analytics [1]. Though, many service providers and vendors now provide advanced data processing capabilities, networks and latest technologies to such organizations via the cloud.

The Big Data will change many fields in future, like business, scientific researches, and so on. Big Data is an assortment of very huge and highly complex data sets. These data sets have a great diversity of types. Hence, it is very difficult to process by using traditional data processing applications. Gartner gave an elaborated definition for Big Data: ‘‘It is a high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization’’. In general, if a data set is more difficult to capture, analysis, and visualize, then such a data set can be called as Big Data.

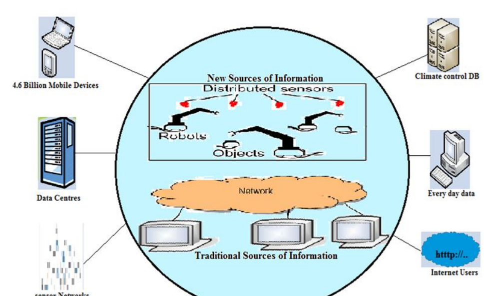

Fig. 1. The future continues to bring new data sources and new volumes of data

The social networking sites, GPS, mobile devices and sensor networks generate new data every day (Figure 1). New kinds of data are born every second, so there are tremendous opportunities to learn, and combine them with existing data and craft new perceptions. However, the challenges arise in capturing data analyzing data and visualizing data. Big Data has changed the ways in doing businesses, managements and researches. In this paper, the survey on various approaches to Big Data analytics, technologies used and how they help to build analytics solutions for Cloud Computing environment have been discussed. It aims to:

- Coordinate and manage data across the services scale from generators of data to information consumers looking for decision making
- Coordinate data collection from various participants and analyze that data to optimize service delivery via Clouds.
- Simplify data management activities to enable positive outcomes for all relevant participants and
- Invest in people, processes, and technology that maximize the value of the organization and enhance its performance.

In addition, this paper suggests a set of recommendations for the research community on future directions on Cloud sustained Big Data computing. In Section II, several Big Data application environments are discussed. The technical issues in Big Data analytics will be introduced in Section III. Big Data challenges are represented in Section IV. A number of principles for designing effective Big Data systems are listed in Section V to handle data-intensive applications. Real-time Data Analytics – example was discussed in Section VI. Section VII describes future trends of Big Data. Section and VIII concludes the paper with future research direction in Big Data.

II BIG DATA APPLICATION ENVIRONMENTS

A. Big Data in Businesses

Handling big data requires a lot of computing power, storage capacity and high speed networks. Cloud Computing makes big data analytics possible for small businesses by authorizing service provides handle the analytics part. Customers tend to use cheaper storages to collect large amounts of data, and then send them to service providers via widely available high-bandwidth networks. However, services based on Clouds become the more efficient path since they're scalable and relatively inexpensive, and they don't require an organization to have employees with data analytics-related skills.

The recent estimates suggested that the volume of business data worldwide doubles every 1.2 years [2]. Taking advantage of sophisticated machine learning techniques to exploit the knowledge hidden in this huge volume of data, they successfully improve efficiency of their pricing strategies and advertising campaigns.

B. Big Data in Public Administration

Public administration also involves Big Data problems [1]. The population of growing countries keeps on increasing. People need different public services for different purposes. For examples, the elder persons require higher level of health care than younger one. Public generates a lot of data in each sector, hence, the volume of data about public administration is extremely huge. For instance, there are almost 3terabytes of data collected by the US Library of Congress. Governments are facing adverse conditions to improve their nation's productivity. They need to provide a high level of public services with significant budgetary constraints. Therefore, they should take Big Data as a potential budget resource and develop tools to get alternative solutions to decrease big budget deficits and reduce national debt levels.The Big Data functionalities for public administration, such as reserving informative patterns and knowledge, improve productivity and higher levels of efficiency and effectiveness.

C. Big Data in Scientific Areas

Due to the development of Computer Sciences, many fields in science become highly data driven areas. Astronomy, social computing, bioinformatics meteorology, and computational biology are greatly producing large volume of data with various types. Astronomers use advanced analysis methods to investigate the origins of the universe. The Large Hadron Collider (LHC) is a particle accelerator that generates 60 terabytes of data per day [2]. The supercomputing clusters in the Center of Climate Simulation (NCCS) generates 32 petabytes of data on climate observations All these scientific areas generate enormous amount of data sets which require automated analysis in making decisions and predictions.

D. Background and Methodology

Knowledge Discovery in Data (KDD) [3] aims to extract irrelevant information from useful one using detailed analysis and interpretation. Data mining [4] targets to learn previously unknown interrelations among unrelated attributes of data sets by applying methods such as machine learning, database systems, and statistics.

Analytics comprises techniques of KDD, data mining, statistical and quantitative analysis, and advanced interactive visualization methods to make decisions and drive actions [5]. Figure 2 depict the common phases of a traditional analytics workflow for Big Data. Data from various sources, such as databases, data streams, data marts, and data warehouses, are used to build data models. The large volume of heterogeneous data demands pre-processing tasks for integrating the data, cleaning it, and filtering it. The preprocessed data is used to train a model and to estimate its parameters. After estimating the model, it is validated and then consumed. Finally, the model is applied to data set. A method called model scoring is used to generate predictions, prescriptions, and recommendations. The results are interpreted and evaluated, and may be used to generate new models.

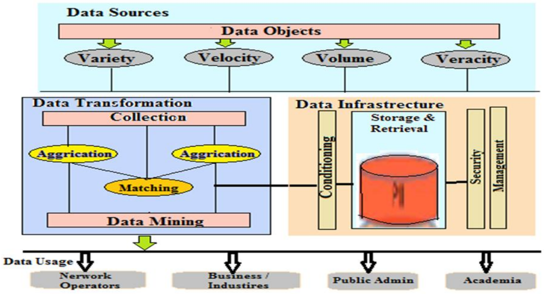

Fig. 2 Workflow in Big Data Analytics

Analytics solutions can be classified as descriptive, predictive, or prescriptive as illustrated in Fig. 2. Historical data are used n descriptive analytics to identify patterns and create reports; it models past behaviour. Predictive analytics tries to predict the future by analyzing current and historical data. Prescriptive solutions help in decision making by determining actions and assessing their impact on business objectives, requirements, and constraints.

Cloud computing provides a useful model for analytics, in which solutions can be presented on the Cloud and used by customers. Several technical issues must be addressed to implement the model, such as data management, tuning of models, privacy, data quality, and data currency. This work highlights technical issues and surveys to provide analytics capabilities for Big Data on the Cloud and consider security is an extensive matter.

III TECHNICAL ISSUES IN BIG DATA ANALYTICS

D. Data Administration

Data preparation is one of the most time-consuming and labor-intensive tasks of analytics. Efficient methods required to perform analytics on large volumes of data that store, filter, transform, and retrieve the data. Deploying data management solutions on Cloud environment involve some challenges as presented in [6], and implementing solutions for analytics on the Cloud face similar challenges. Solutions for Cloud analytics need to contemplate the several Cloud deployment models adopted by enterprises. The following scenarios are considered regarding the availability of data and analytics models [7]:

(i) Both data and models are private;
(ii) Data is public, models are private;
(iii) Both data and models are public; and
(iv) Data is private, models are public.

Jensen et al. [8] supports on deployment models for Cloud analytics solutions using privately hosted software and infrastructure, private analytics hosted on third party infrastructure, and public model where the solutions are hosted on a public Cloud. To attain economies of scale and elasticity, Cloud-enabled Big Data analytics needs to discover ways to allocate and utilize specialized resources in appropriate manner.

E. The four V's of Big Data

Big Data is characterized by variety, velocity, volume and veracity. Variety represents the data types, velocity refers to the rate at which the data is produced and processed, and volume expresses the amount of data. Veracity refers to how much the data can be trusted given the reliability of its source [9], whereas value corresponds to the economical worth that an organization derives from employing Big Data computing solutions. Regarding Variety, more volume of data has been made publicly available for scientific and business practices. Eco-

Intelligence [10] platform is an example for illustrating the need of variety within a single analytics application. It was designed to analyze large volumes of data to support urban area planning and development. It aims to control and process data from several sources, such as sensors, web sites, television and radio, and helps in urban development.

F. Data storage

To store and retrieve large volume data required by Big Data, many solutions were proposed and currently used in Clouds. For instance, Internet-scale file systems such as the Google File System (GFS) [7] attempts to provide reliable, robust and scalable solutions for data storage in Clouds. Other systems provide solutions for object-store capabilities in which files are replicated across several geographical areas to increase data readiness, redundancy, and scalability. The main aspect in providing performance for Big Data analytics applications is the data locality. In the Big Data analytics context, MapReduce provides an ideal model for data locality to improve the performance of applications. An open source solution to implement MapReduce is Hadoop, which allows clusters creation among data sets that uses the Hadoop Distributed File System (HDFS) to partition and replicate datasets across nodes where they are consumed by mappers. HDFS reduces the effect of failures by replicating datasets across a number of nodes. Thusoo et al. [11] used Hadoop to develop an analytics platform to process the large volume of Face-book's data sets. The drawback of MapReduce implementations is that they require the users to learn a new set of APIs to construct analytics solutions for the Cloud. The comparison between the Parallel Virtual File System (PVFS) and the HDFS was done by Hannes Muhleisen et al. [12], expressed that the PVFS did not extant significant improvement in completion time and throughput compared to the

HDFS.

G. Data integration

Data warehouses based on Clouds create problems with respect to data integration and the addition of new data sources. Standard formats and interfaces are necessary to meet the needs of a large number of consumers [13]. Composite spaces and space inheritance integrates data from one or more parent spaces with additional data added to the composite space and provides a Software as a Service (SaaS) solution that offers analytics methods on a payment model. The tool IVOCA aimed at Customer Relationship Management (CRM) that consists of both structured and unstructured data and provides text mining, data linking, and classification facilities to smooth analyst's tasks and reduce the time to insight. The Business Process Execution Language (BPEL) is used to exchange data by passing references to data between services to reduce the execution time and guarantee the correct data processing of an analytics process. A generic data Cloud layer is introduced to handle heterogeneous data Clouds, and is responsible for mapping generic operations to each Cloud implementation.

H. Data processing and resource management

Amazon EMR [14] allows customers to create Hadoop clusters to process large volume of data using the Amazon Elastic Compute Cloud (EC2) and uses Amazon Web Services for data storage and transfer. The HDFS file system in Hadoop is used to partition and replicate datasets across multiple sites and a mapper is used to access local data stored on the cluster node. Hadoop provides data parallelism and fault tolerance; however, it takes more time to load data into HDFS and the mapper lack in reuse of data. A data analytics system called Starfish built on Hadoop aims on improving the performance of clusters without requiring users to understand the available configuration options. Starfish uses several techniques at various stages to optimize the execution of MapReduce tasks. It optimizes workflows by minimizing the influence of data imbalance and by load balancing for executions. Elastisizer of the Starfish automates the decision making process by integrating simulation and model-based estimation to address various issues on workload performance. A MapReduce runtime for Windows Azure called Daytona controls the storage services provided by Cloud infrastructure of Azure as the source and destination of data. It uses features of the Clouds to provide load balancing and fault tolerance. Daytona is a master–slave architecture in which the master is responsible for job scheduling and the slaves for performing map and reduces operations.

IV BIG DATA CHALLENGES

The challenges in handling Big Data problems lies in data capture, storage, searching, sharing, analysis, and visualization. The main challenge occurs in the nature of computer architecture exists for several years which are CPU-centric but I/O-poor. This imbalance of the system still limits the development of new things from Big Data. As per the Moore's Law, the performance of CPU doubles every one and half year, and the performance of storage drives is also doubling at the same rate. However, the access time of the storage device has slightly improved over the last decade. Moreover, information is increasing at exponential rate simultaneously, but the improvement in information processing methods still relatively slower. In Big Data applications, the contemporary techniques and technologies cannot preferably solve the real problems, especially for real-time analysis. Data inconsistence and data incompleteness, scalability, timeliness and data security are the challenges in Big Data analysis [8]. As the previous step to data analysis, data must be well-constructed. However, the heterogeneity of data sets in Big Data problems enables us to drive efficient data representation, access, and unstructured or semi-structured data analysis in the further researches. Current databases are severely susceptible to inconsistent, incomplete, and noisy data due to very large sizes of data and their origin from heterogeneous sources,. Therefore, a number of data preprocessing techniques, such as data cleaning, data integration, data transformation and date reduction, can be employed to eliminate noise and correct discrepancies.

A. Data capture and storage capacity

Data are being gathered from different resources such as mobile devices, software logs, cameras, microphones, readers, wireless sensor networks, and so on and data sets grow in size. Every day, there is 2:5 quintillion bytes of data created and this number keeps on growing exponentially. The capacity to store information has approximately doubled about every 3 years. In many fields, valuable data is often being deleted just because of no enough space to store it. These valuable data are created and captured at high cost, but ignored finally.

The way of capturing and storing Big Data had changed with data storage device, data storage architecture, and data access mechanism [15]. More storage mediums and higher I/O speed are required to meet the challenges. The current storage mechanisms cannot provide same performance for both the sequential and random I/O devices simultaneously, which requires designing new ways of storage subsystems for Big Data processing systems. Data-access platforms, such as CASTOR, GPFS, dCache and Scalla are employed to exhibit the large scale validation and performance measurements. Data search schemes also lead to high overhead and latency [16].

B. Data transmission

The network bandwidth is the major bottleneck in the

Cloud data storage, especially when the volume of data is large. On the other side, Cloud data storage leads to data security problems as the requirements of data integrity checking. Many security schemes were developed and proposed under different security models [17].

C. Data curation

Data curation is the process of finding data, reuse, retrieval, quality assurance, value addition, and data preservation overtime which include authentication, archiving, management, preservation, retrieval, and representation as sub-fields. The traditional database management systems are unable to process Big Data due to its high complexity and large voluminous. The classical approach for structured data management has two parts, one is a schema to storage the data set, and another is a relational database for data retrieval. The two traditional approaches for managing large-scale datasets in a structured way are the data warehouses and data marts.

‘‘Not Only SQL’’ (NoSQL) database [18], is used for large and distributed data management. The Big Data platforms adopt NoSQL to break and surpass the rigidity of normalized RDBMS schemas. For instance, Hbase is one of the most famous used NoSQL databases. Other NoSQL implementations include Google BigTable, Apache Hadoop, MapReduce, and SimpleDB that use NoSQL.

D. Data analysis

The biggest challenge in Big data is scalability when deal with the Big Data analysis tasks. In the Big Data analytical techniques, increment algorithms possess good scalability property. For real-time Big Data applications such as social networks, finance, biomedicine, intelligent vehicular networks, and internet of thing, the timeliness is at the top priority. Data security problems such as data protection, intellectual property protection, privacy protection, commercial secrets and financial information protection are dealt with more priority [19]. For applications of Big Data, problems due to security are more obstinate for several reasons. First, the size of Big Data is extremely large, channeling the protection approaches. Second, it leads to heavier workload of the security. Most Big Data are stored in a distributed manner, and many threats from networks aggravate the problems.

E. Data visualization

Data visualization represents the knowledge in more instinctive and effective manner by using different graphs. Data visualization is used to convey information more easily by providing knowledge concealed in the complex and large-scale data sets. Big Data visualization tool like Tableau used to transform large volume of data sets into interactive pictures. Users visualize search relevance and quality to monitor the feedback from customers and conduct sentiment analysis. Data Visualization process is difficult for Big Data applications due to the large size and high dimension of Big Data. Many Big Data visualization tools lack in functionalities, scalability and response time. New frameworks are necessary for characterizing the evolution of the uncertainty information through analytical processes. The deficiency of ability to capture values from Big Data will be a significant constraint [20].

V GUIDELINES FOR DESIGNING BIG DATA SYSTEM

It is evident that the Big Data analytics are more complex than the traditional data analysis systems. Development of new technologies and new way of thinking about data are necessary to exploit Big Data. Some necessary principles in designing Big Data Analytics systems are outlined here [21] to guide the development of Big Data Systems.

1. Top Priority can be given to good architectures and frameworks.
2. Support a variety of analytical methods
3. No size fits all
4. Bring analysis towards data
5. In-Memory processes must be distributable
6. Data storage must be distributable for in-memory storage.
7. Coordination is needed between processing and data units.

VI REAL-TIME DATA ANALYTICS – AN EXAMPLE

A. Manufacturing System

The appraisal of diagnostics and prognostics for real-time development of dynamic production system effectiveness requires data analytics solutions based on open standard protocols and standards. Data analytics solutions for manufacturing system need data management in manufacturing that include data acquisition from different manufacturing stages, data processing, data visualization, and semantics. The existing solutions for data analytics are based on proprietary models. Also solutions available in the market are mostly targeted towards original equipment manufacturers (OEMs) and seldom available as reconfigurable open applications suited for small and medium scale enterprises (SMEs). Data analytics infrastructures developed for manufacturing System includes open standard protocols, predictive analytical models, and standards, deliver a framework for real-time improvement of dynamic production system efficiency. The project developed for manufacturing system will enable deployment of reconfigurable and cost-effective open platform for manufacturing data analytics based on open standards and protocols, suited for both large OEMs and SMEs.

B. Objective:

Develop methods, standards, and protocols for data analytics to facilitate real-time diagnostics and prognostics that will extensively increase the efficiency of dynamic production systems

C. The Research Activities

The research activities include the following:
Analyze requirements and the state of the art in management of manufacturing data that includes:
Data acquisition, data modeling, input validation, and data management Distributed data applications characterized by continuous analysis and latency requirements A schema for the predictive modeling can be developed that includes optimization models based on the Sustainable Process Analytics Framework developed under Sustainable Manufacturing Program.
Explore probabilistic graphical models such as dynamic Bayesian network that have shown much efficiency in the integration of uncertainty information across complex production networks. Identifying standards to signify many predictive modeling techniques to fully support data analytics for manufacturing applications that include:

a. Analysis of features, limitations, and applications of the predictive modeling technique(s).
b. Extending existing standards for manufacturing applications including semantics.
c. Identification of manufacturing data (product, process, and asset) for analytics.

VI FUTURE TRENDS OF BIG DATA

Techniques in Big Data analytics should kindle the development of new data analytic tools and algorithms that should facilitate scalable, accessible, and sustainable data infrastructure and increase human and social interactions. A paradigm shift in scientific research will be well-established in future owing to novel mathematical models, statistical techniques, new data mining tools and advanced machine learning algorithms. They focus on common research interests in Big Data analytics across the world. Several ongoing or underlying techniques and technologies to harness Big Data are discussed in the following sub-sections.

A. Granular computing

Granular computing (GrC) [22] is a general computation model for successfully implementing granules such as classes, clusters, subsets, groups and intervals to build an effective computational model for complex applications with huge amounts of data, information and knowledge. It reduces the size of data into different level of granularity and provides a significant transform from the current machine-centric approach to human-centric. Set theory, fuzzy sets, rough sets and random sets linked together in a comprehensive manner in GrC. The information represented by different level of granules in GrC highlights distinct knowledge, features, and patterns, hides irrelevant features and shows valuable ones. Hence, GrC is a very useful platform to design more effective machine learning and data mining algorithms. The usage of GrC techniques in Big Data applications depend on the confidence and accuracy of results the system required.

B. Cloud computing

The growth of virtualization technologies have made supercomputing more accessible and reasonable. The use of virtual machines leads to the development of cloud computing [23] as shown in Figure 3, which is one of the BigData techniques. The combination of virtual machines and large numbers of processors leverage internet-based companies to invest in large-scale computational clusters and advanced data-storage systems. Cloud computing delivers applications and many services over the Internet. It leads to the utility computing, i.e., pay-as-you-go computing. Big Data forms a framework for conferring cloud computing options. Another advantage of cloud environments is the use of cloud storage shown in Figure 4, which provides a tool for storing Big Data.

Fig. 3. Cloud Computing – Logical Diagram

Fig. 4. Cloud Storage

A frame work called CloudView [23] is used for storage, processing and analysis of massive machine maintenance data in a cloud computing environment. Cloud computing addresses one of the challenges relating to transferring and sharing data, because data sets and analysis results held in the cloud can be shared with others. The Big Data problems will drive the cloud computing to a high level of development.

C. Bio-inspired computing

Human brain manages thousand tera-bytes of data every day. The biological computing system of our brain works well in a distinct way compared to computers. Biological computing models are good for Big Data because they have mechanisms to organize, access and process data in the same ways our brain deal with every day. Computational intelligence has methodologies and approaches to address complex real-world problems. Biologically inspired Computing is shown in Figure 5, provides tools to solve Big Data problems that range from hardware design to software design. It can be classified as phylogeny, ontogeny, and epigenesis [24]. It employs bio-inspired mechanisms to find the optimal solution for data service by considering the cost of data management and service maintenance. Biological molecules such as DNA and proteins are used to develop Bio-computers to conduct computational calculations involving storing, retrieving, and processing data. The important feature of bio-computer is that it integrates biologically derived resources to do computational functions and receive intelligent performance. Bio-inspired technologies will provide a large amount of funds and human resources in future related research activities.

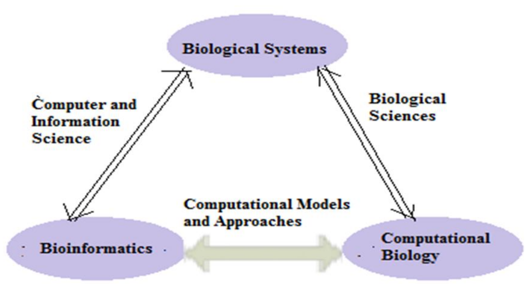

Fig. 5 Bio-inspired Computing Platform

D. Quantum computing

A quantum computer has very large memory which is larger than its deceptive physical size. It can manipulate very large set of inputs simultaneously. Today's Big Data problems can be efficiently solved by quantum computers. For example, D-Wave Systems developed a quantum computer, called ''D-Wave one'' with 128 qubits processor and ''D-Wave two'' with 512 qubits. In essence, quantum computing [25] is used to exploit the powerful laws of quantum mechanics to process information.

Quantum computing can exploit on the concepts of ''superposition'' and ''entanglement''. Many big problems can be solved by quantum computers compared with classical computers because of quantum algorithms, such as Simon's algorithm, Shor's algorithm, and other algorithms for simulating quantum systems, are more efficient and faster than traditional one. Many universities, institutes, national governments and military funding research groups are working on quantum computing to develop quantum computers for both civilian and national security purposes.

VII CONCLUSION

The trends in Big Data are being viewed by industries as a way of gaining advantage over their competitors. However, the Big Data analytics is still a powerful, more challenging and time consuming task that requires expensive software, large computational infrastructure, and effort. Cloud Computing helps in overcomes from these problems by providing resources on-demand with costs proportional to the actual usage. Furthermore, it allows infrastructures scaled up well and adapting the system to the actual demand. Although Cloud offers flexible capacity to supply computational resources on demand, the area of Cloud sustained Big Data analytics is still in its early days. In this paper, the aspects of analytics workflows are discussed, and the state-of-the-art of each phase in the context of Cloud- was surveyed. For each phase, the enduring work was studied and key challenges were discussed. This survey concluded with an analysis of business models for Cloud-assisted data analytics.

The future work can include:

(i) The Standards and APIs enable users to easily switch among solutions

(ii) The ability of receiving the most of the capacity of the Cloud infrastructure which enable users to describe the problem in simple terms and decomposing such high-level description in highly parallel subtasks and keeping efficiency even for large numbers of computing resources. The only limitations will be the market issues, namely the relation between the cost for running the analytics and the financial return brought for the obtained knowledge.

REFERENCES

[1] Neal LeavittBringing, "Big Analytics to the Masses, Technology News" by IEEE Computer Society, 0018-9162/13/2013.

[2] James Manyika, Michael Chui, et. al, "Big Data: The next Frontier for Innovation, Competition, and Productivity", McKency Global Institute, 2012.

[3] Itamar Arel, et.al., "Deep Machine Learning – a new frontier in Artificial Intelligence Research", IEEE Computin Intelligence (2010) 13 -18.

[4] Dawei Jiang, Antony K.H. Tung, Gang Chen, "Map-join-Reduce: towards scalable and efficient data analysis on large clusters", IEEE Transacrion on Knowledge and Data Engineering, (23) (2011) 1299-1311.

[5] Trvor Hastle, et.al., "The Elements of Statistical Learning, Data Mining Inference and Prediction", 2nd Edition, Springer, 2009.

[6] Mohsen Jamali, Hassan Abolhassani, "Different aspects of Social network Analysis" in IEEE/ACM International Conference on Web Intelligence, 2006, pp 66 -72.

[7] Vamsee Kasavajhala, "Solid state device vs Hard disk drive price and performance study", Dell Power Vault Tech, Mark (2012).

[8] R.P. Ishii, R.F. de Mello, "A History-based heuristic to optimize data access in distributed environments", IASTED International Conference on Parallel and Distributed Computing Systems, 2009.

[9] Tamer M. Ozsu, Patric Valdurier, "Principles of Distributed data base System", 3rd Edition, Springer, 2011.

[10] Andrew Pavlo, et. al., "A Comparison of approaches to large-scal data analysis in SIGMOD 2009", Proc. Of ACM Intel. Conference on Management of Data, 2000, pp 165 – 178.

[11] Tadashi Nakno, "Biological Computing based on living cells and cells communication in 2010", 3th Intel Conference on Network-based Information System (NBiS) 2010, pp 42 – 47.

[12] Hannes Muhleisen, Kathrin Dentler, "Large-scale storage and reasoning for scemantic data using swarms", IEEE Computing Intelligence Magazine, 7(2) 2012 pp 32-44.

[13] M. Chen, S. Mao, and Y. Liu, "Big data: a survey," Mobile Networks and Applications, vol. 19, no. 2, pp. 171–209, 2014.

[14] Pentaho Business Analytics, 2012, http://pentaho.com/explore/ pentaho-business-analytics.

[15] J. Mervis, "Agencies rally to tackle big data," Science, vol. 336, no. 6077, p. 22, 2012.

[16] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: challenges, issues, and opportunities," in Database Systems for Advanced Applications, pp. 1–15, Springer, Berlin, Germany, 2013.

[17] J. Manyika, C. Michael, B. Brown et al., "Big data: The next frontier for innovation, competition, and productivity," Tech. Rep., Mc Kinsey, May 2011.

[18] www.nosql-database.org/

[19] R. T. Kouzes, G. A. Anderson, S. T. Elbert, I. Gorton, and D. K. Gracio,"The changing paradigm of data-intensive computing," IEEE Computer, vol. 42, no. 1, pp. 26–34, 2009.

[20] V. Mayer-Schönberger and K. Cukier," Big Data: A Revolution That Will Transform How We Live, Work, and Think", Eamon Dolan/Houghton Mifflin Harcourt, 2013.

[21] Nathan Marz, James Warren, "Big-Data: Principles and best practices of scalable real-time data systems", Manning, 2012.

[22] Witold Pedrycz, Andrzej, Vladik Kreinovich, "Hand book of Granular Computing", WILEY, 2008.

[23] D. Wang, "An efficient cloud storage model for heterogeneous cloud infrastructures," Procedia Engineering, vol. 23, pp. 510–515, 2011.

[24] Eduardo Sanchez, "Phylogeny, Ontogeny, and Epigenesis: Three Sources of Biological Inspiration for Softening Hardware", Springer, 1996

[25] www.dwavesys.com/quantum-computing/