© 2015 IJSRST | Volume 1 | Issue 5 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Strategies for Revealing and Understanding Complex Relationships towards Big Data Processing Frameworks

Avanthi Nagelli[1], Dr. Chandra Shekar[2]
[1]Sr. Software Engineer, People Tech Group, India
[2]Professor, Avanthi Engineering College, Kodhada, India

ABSTRACT

The term 'Big Data' was first coined in 1998 by John Mashey during a Silicon Graphics (SGI) slide deck presentation titled "Big Information in addition to the Observing Wave of InfraStress". Big Data mining refers to the process of extracting valuable insights from large datasets or streams of data that were previously impractical to analyze due to their size, complexity, and cost. The Big Data challenge is becoming one of the most exciting prospects for the coming years. In this article, we provide a comprehensive overview of the topic, its current state, issues, and future prospects. We evaluate the Big Data processing framework for complex and evolving relationships.

Index Terms : Data Mining, Big Data, Processing Framework

I. INTRODUCTION

Big data mining was deemed necessary in its early stages, with the first publication discussing 'big data' appearing in a record exploration magazine in 1998 by Weiss and Indrukya. The first academic paper featuring the term 'big data' in its title appeared a little later in 2000 in a paper by Diebold.

The term 'big data' originated from the fact that we generate a massive amount of data every day. Usama Fayyad, in his talk at the KDD Big Mine ‘‘12 Workshop, shared staggering data about internet usage, which is as follows: Google receives over one billion queries daily, Twitter has more than 250 million tweets every day, Facebook generates more than 800 million updates per day, and YouTube has more than 4 billion views per day. The data generated today is estimated to be in the order of zettabytes, and it is growing at a rate of about 40% each year. A new significant source of data is going to be generated from mobile phones, and major corporations such as Google, Apple, Facebook, and Yahoo are beginning to delve into this data to identify useful patterns to enhance user experience.

'Big data' is a ubiquitous term that still confuses many. It has been used to describe all kinds of concepts, including vast amounts of data, social media analytics, next-generation data management capabilities, real-time data, and more. Nevertheless, organizations are beginning to understand and learn how to process and analyze a significant range of data in new ways. In doing so, a small but rapidly expanding group of pioneers is achieving innovative business outcomes. Across various industries worldwide, executives acknowledge the need for more information about how to handle big data. However, despite the seemingly relentless media attention, it can be challenging to find comprehensive information on what businesses are genuinely doing. Therefore, we sought to

gain a better understanding of how companies perceive significant data and to what extent they are currently using it to benefit their organizations.

II. BIG DATA CHARACTERISTICS: HACE THEOREM

Big Data is characterized by large-volume, heterogeneous, independent resources with distributed and decentralized control. The challenge with this data is that it is difficult to extract useful insights from it. Imagine a group of blind men attempting to describe an elephant by only feeling a part of it. Each of them will provide a different description based on the information they have gathered. Similarly, Big Data is like the elephant, and the blind men represent the different sources of data. To make it even more complex, the data is constantly changing and evolving, and the sources may have privacy concerns or speak different languages.

One of the key features of Big Data is the massive amount of data represented by various and different dimensionalities. This is because different data collectors use their own schemata for data recording, and different applications also produce different data. For example, in the medical field, each person can be represented by demographic information, X-ray or CT scans, or DNA data. The diversity and varied dimensionality of the data become significant issues when trying to integrate data from all sources.

Another characteristic of Big Data is the autonomous resources with distributed and decentralized controls. Each resource can generate and collect data without relying on centralized control, similar to how each web server operates independently on the Internet. However, the large volumes of data also make the system vulnerable to attacks or failures if there is no centralized control. To ensure consistent services and fast responses, companies like Google, Facebook, and Walmart deploy multiple server farms around the world.

In short, Big Data is a complex and challenging field that requires expertise in data integration, analysis, and security.

III. COMPLEX AND EVOLVING RELATIONSHIPS

As the amount of Big Data increases, so does the complexity and relationships within the data. In traditional data systems, the focus is on finding the best quality values to represent each individual. This is done by using data fields such as age, gender, income, education history, etc. to characterize everyone. However, this sample-feature representation treats each person as an independent entity without considering their social links, which is one of the most important factors of human society. People form friend groups based on their shared interests or biological relationships, and these social links are not only present in our daily lives but also in online worlds. Social media platforms like Facebook and Twitter are distinguished by features such as friend-relationships and followers. The connections between people make the entire data representation and decision-making process more complex. In the sample-feature representation, individuals are considered similar if they share similar quality values, while in the sample-feature-relationship representation, two people can be grouped together based on their social links even if they don't share anything in common in terms of their characteristics. In a dynamic world, the characteristics used to represent people and the social ties used to represent our relationships may also evolve over time and space. This is a challenge in Big Data applications, where the goal is to take into account the non-linear, many-to-many data connections and their changing nature to identify useful patterns in Big Data collections.

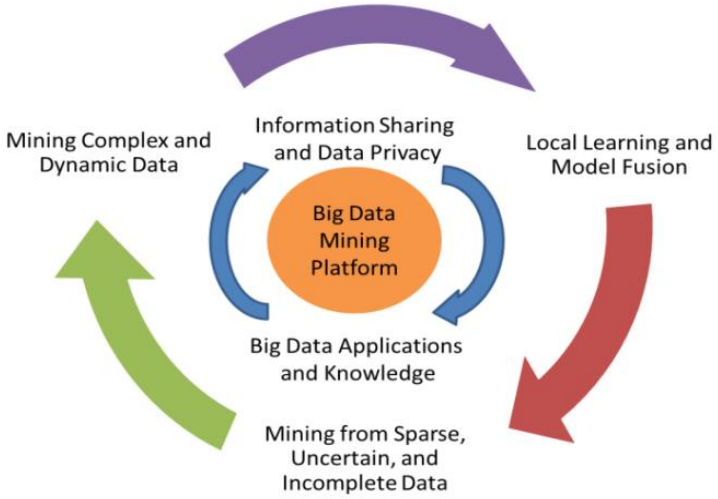

Figure 1 : A Big Data processing framework

The analysis study identifies three tiers of obstacles and centers around a "Big Data exploration tool" (Price I), which focuses on accessing and processing low-level data. Challenges on information sharing, privacy, and understanding Big Data application domains fall under the second tier of issues (Price II), which focuses on high-level semantics, processing domain knowledge, and privacy concerns. The outermost tier deals with Price III issues related to real exploration algorithms.

IV. DATA MINING FOR BIG DATA

Data mining, also known as knowledge or information discovery, is the process of analyzing data from various perspectives and consolidating it into useful information that can be used to improve revenue, reduce costs, or both. Technically, data mining involves finding relationships or patterns among large amounts of data in a relational database. Data mining includes six tasks or functions: classification, estimation, prediction, business rules, clustering, and summary.

Classification is the process of categorizing data according to specific criteria. The most common classification algorithms in data mining are decision trees, k-nearest neighbor classifiers, naive Bayes, Apriori, and AdaBoost. The classification process involves examining the attributes of a newly presented object and assigning it to a predefined class.

Estimation deals with continuously valued outcomes. Given some input data, we use estimation to come up with a value for some unknown continuous variables, such as revenue, height, or credit card balance.

Prediction is a statement about how things will happen in the future, often but not always based on experience or knowledge. Prediction can be a statement in which some result is anticipated.

Business rules are guidelines that indicate specific relationships among a set of items (such as "occur together" or "one suggests the other") in a database.

Clustering can be considered the most important unsupervised learning problem, and like every other problem of this type, it deals with finding a structure in a collection of unlabeled data.

Big data	Data mining
Big data is a term for large data set.	Data mining refers to the activity of going through big data set to look for relevant information
Big data is the asset	Data mining is the handler which provide beneficial result.
Big data" varies depending on the capabilities of the organization managing the set, and on the capabilities of the applications that are traditionally used to process and analyze the data.	Data mining refers to the operation that involve relatively sophisticated search operation

Table 1 : Difference between Data Mining and Big Data

V. CONTROVERSY ABOUT BIG DATA

Big Data is a topic that has been widely discussed. Here's what you need to know:

There's no need to differentiate between Big Data analytics and data analytics because the amount of data will only increase in the future, and it will never be too small to analyze.
Hadoop-based computing systems are often marketed as the best tools for Big Data management, but that's not always the case. For instance, MapReduce may not be the best programming platform for medium-sized businesses.
In real-time analytics, the recency of data is more important than its size.
Claims of accuracy can be misleading because when the number of variables increases, so do the number of fake correlations. For example, a study found that the S&P 500 stock index was linked to butter production in Bangladesh and other unrelated factors.
Larger data sets do not necessarily mean better data. The quality of data matters more than its size, and it should be representative of what we are looking for.
There are ethical concerns about accessing and analyzing people's data without their knowledge or consent.
The lack of access to Big Data creates new digital divides between individuals and companies who can analyze it and those who cannot. Access to Big Data can provide organizations with a competitive advantage that others may not have.

VI. CONCLUSION

It is common for real-world systems to undergo development over time. For instance, a doctor's treatment plans may need to be adjusted based on various factors like the patient's family economic status, medical insurance, treatment outcomes, and changes in chronic diseases. In the process of exploring new knowledge, researchers set goals to evaluate the impact of proposed changes and fundamental modifications due to changing conditions

in data streams. This paper examines the major data processing platform used for complex and evolving relationships.

REFERENCES

[1]. Aral S. and Pedestrian D. 2010, Recognizing prominent as well as vulnerable participants of socials media, Science, vol.337, pp.337-341.

[2]. Machanavajjhala as well as Reiter 2012, Ashwin Machanavajjhala, Jerome P. Reiter: Big privacy: safeguarding privacy in big data. ACM Crossroads, 19(1): 20-23, 2011.

[3]. Banerjee and also Agarwal, Soumya Banerjee, Nitin Agarwal, Analyzing cumulative actions from blog sites making use of swarm intelligence, Understanding as well as Info Equipment, December 2011, Volume 33, Issue 3, pp 523-547

[4]. Peddyreddy. Swathi, "Approaches And Objectives towards Financial Management", International Journal of Advanced in Management, Technology and Engineering Sciences, Volume IV, Issue I, 2014

[5]. Peddyreddy. Swathi, "An Overview On The Types Of Capitalization", International Journal of Advanced in Management, Technology and Engineering Sciences, Volume VI, Issue I, 2016

[6]. Peddyreddy. Swathi, "Architecture And Editions of Sql Server", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Volume 2, Issue 4, May-June-2017

[7]. Peddyreddy. Swathi, "Scope of Financial Management and Functions of Finance", International Journal of Advanced in Management, Technology and Engineering Sciences, Volume III, Issue 1, 2013

[8]. Peddyreddy. Swathi, "A Study On Security Towards Sql Server Database", JASC: Journal of Applied Science and Computation, Volume V, Issue II, February 2018

[9]. Peddyreddy. Swathi, "A Comprehensive Review on The Sources of Finance", International Journal of Scientific Research in Science, Engineering and Technology, Volume 1, Issue 4, July-August 2015

[10]. Peddyreddy. Swathi, "A Study on SQL - RDBMS Concepts And Database Normalization", JASC: Journal of Applied Science and Computations, Volume VII, Issue VIII, August 2020

[11]. Peddyreddy. Swathi, "A Comprehensive Review on SQL - RDBMS Databases", Journal of Emerging Technologies and Innovative Research, Volume 6, Issue 3, March 2019.

[12]. Peddyreddy. Swathi, "An Overview on the techniques of Financial Statement Analysis", Journal of Emerging Technologies and Innovative Research, Volume 1, Issue 6, November 2014

[13]. Peddyreddy. Swathi, "COMPLEXITY OF THE DBMS ENVIRONMENT AND REPUTATION OF THE DBMS VENDOR", Journal of Interdisciplinary Cycle Research, 13 (3), 2054-2058

[14]. Peddyreddy. Swathi, "Implementation of AI-Driven Applications towards Cybersecurity", JASC: Journal of Applied Science and Computations, 7(8), 127-131

[15]. Peddyreddy. Swathi. (2022). Implications For Research In Artificial Intelligence. Journal of Electronics,Computer Networking and Applied Mathematics(JECNAM) ISSN : 2799-1156, 2(02), 25–28. Retrieved from http://journal.hmjournals.com/index.php/JECNAM/article/view/447

[16]. Peddyreddy. Swathi. (2022). A Study On The Restrictions Of Deep Learning. Journal of Artificial Intelligence,Machine Learning and Neural Network (JAIMLNN) ISSN: 2799-1172, 2(02), 57–61. Retrieved from http://journal.hmjournals.com/index.php/JAIMLNN/article/view/444

[17]. Peddyreddy. Swathi. (2022). Industry Applications of Augmented Reality and Virtual Reality. Journal of Environmental Impact and Management Policy(JEIMP) ISSN:2799-113X, 2(02), 7–11. Retrieved from http://journal.hmjournals.com/index.php/JEIMP/article/view/453

[18]. Keerthi Vuppula, "Access Control with Energy Restrictions for IoT Nodes Power", "International Journal on Applications in Engineering and Technology", Volume 4, Issue 3: September 2018, pp 1 – 6

[19]. Keerthi Vuppula, "Method for Recognizing Palmprints Using Neighboring Direction Indicator", "International Journal of Scientific Research in Engineering and Management (IJSREM)", Volume 05, Issue: 11, November - 2021

[20]. Keerthi Vuppula, "Smart Door Unlock System Using Face Recognition and machine learning", "JOURNAL FOR INNOVATIVE DEVELOPMENT IN PHARMACEUTICAL AND TECHNICAL SCIENCE", Volume-2, Issue-3 (Mar-2019)

[21]. Keerthi Vuppula, Dr. Narsimha Reddy, "Computer-Aided Diagnosis for Diseases using Machine Learning", "International Journal of Scientific Research in Engineering and Management (IJSREM)", Volume 04, Issue 12, November - 2020

[22]. Keerthi Vuppula, Dr. K. Mounika Reddy, "Design of Smart Agriculture System Using Internet of things", "International Journal on Applications in Engineering and Technology", Volume 1, Issue 11, November 2015, pp 7 – 12

[23]. Keerthi Vuppula, "An advanced machine learning algorithm for fraud financial transaction detection", "Journal For Innovative Development in Pharmaceutical and Technical Science (JIDPTS)", Volume 4, Issue 9, Sep 2021

[24]. Keerthi Vuppula, Dr. Narsimha Reddy, "Analysis on Supervised machine learning based Flower Classification", "INTERNATIONAL JOURNAL FOR RESEARCH & DEVELOPMENT IN TECHNOLOGY", Volume-15, Issue-2 (Feb-21)

[25]. Keerthi Vuppula, Dr. Narsimha Reddy, "Facial emotion detection using machine learning algorithm K-nearest neighbor", "INTERNATIONAL JOURNAL FOR RESEARCH & DEVELOPMENT IN TECHNOLOGY", Volume-13, Issue-2(Feb-20)

[26]. Keerthi Vuppula, "Internet of things based Smart Watch for Health Monitoring of Elderly People", "International Journal on Applications in Information and Communication Engineering", Volume 5, Issue 1, August 2019 , pp 82 –88

[27]. Keerthi Vuppula, "Design of Internet of things-based human-computer interface system", "International Journal on Applications in Basic and Applied Sciences", Volume 1, Issue 5, December 2013, pp 18-23.