ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

A Study on Effective Business Logic Approach for Big Data Mining

T. Sathis Kumar

Assistant Professor, Dept of C.S.E, Saranathan College of Engineering, Trichy, Tamil Nadu, India.

ABSTRACT: Big data is an magical term that describes a collection of data sets which are large and complex, growing data sets with multiple, autonomous sources , it contain structured and unstructured both type of data. Anyway big data doesn't refer to any specific quantity, since data comes from everywhere, so useful data can be extracted from this big data with the help of data mining. This paper presents a HACE theorem that characterizes the features of the Big data revolution, and proposes a Big data processing model, from the data mining perspective. Data mining is a technique for discovering the patterns as well as descriptive, understandable models from large scale data. Here I analyze the challenging issues and the features of the Big data revolution.

KEYWORDS: Big data, data mining, heterogeneity, HACE theorem, autonomous sources

I. INTRODUCTION

Nowadays we are using the huge source of information via worldwide network. We live in a data driven world, the direct result of advents in information and communication technologies. Moreover, the need of information from these vast amounts of data is even more pressing for enterprises. Mining information from raw data is an extremely vital and tedious process in today's information driven world. Enterprises today rely on a set of automated tools for knowledge discovery to gain business insight and intelligence. In fact every one needs to know how to make use of business logic components. Main tools for getting information from these vast amounts are automated mining tools, specifically speaking data mining, text mining, and web mining. The Big data analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits. Big data can be analysed with the B2B,B2C using the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis. So the biggest problem is that how to handle such unstructured data which comes in business logic manners in real time.

II. BIG DATA CHARACTERISTICS: HACE THEOREM

Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, with complex and evolving relationships among data. These characteristics make it an extreme challenge for discovering useful knowledge from the Big Data. The term Big Data literally concerns about data volumes, HACE theorem suggests that the key characteristics of the Big Data are 1. Huge with heterogeneous and diverse data sources:- One of the fundamental characteristics of the Big Data is the huge volume of data represented by heterogeneous and diverse dimensionalities. This huge volume of data comes from various sites like Twitter, MySpace, Orkut and LinkedIn etc. 2. Decentralized control: - Autonomous data sources with distributed and decentralized controls are a main characteristic of Big Data applications. Being autonomous, each data source is able to generate and collect information without involving (or relying on) any centralized control. This is similar to the World Wide Web (WWW) setting where each web server provides a certain amount of information and each server is able to fully function without necessarily relying on other servers 3. Complex data and knowledge associations: - Multi structure, multi source data is complex data. Such combined characteristics suggest that Big Data require a "big mind" to consolidate data for maximum values.

While the volume of the Big Data increases, so do the complexity and the relationships underneath the data. In an early stage of data centralized information systems, the focus is on finding best feature values to represent each observation. This is similar to using a number of data fields, such as age, gender, income, education background, and so on, to characterize each individual. This type of sample feature representation inherently treats each individual as an independent entity for business without considering their social connections, which is one of the most important factors

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

of the human society. For example, major social network sites, such as Face book or Twitter, are mainly characterized by social functions such as friend-connections and followers (in Twitter). The correlations between individuals inherently complicate the whole data representation and any reasoning process on the data. In the sample-feature representation, individuals are regarded similar if they share similar feature values, whereas in the sample-feature-relationship representation, two individuals can be linked together (through their social connections) even though they might share nothing in common in the feature domains at all. In a dynamic world, the features used to represent the individuals and the social ties used to represent our connections may also evolve with respect to temporal, spatial, and other factors. Such a complication is becoming part of the reality for Big Data applications, where the key is to take the complex (nonlinear, many-to-many) data relationships, along with the evolving changes, into consideration, to discover useful patterns from Big Data collections.

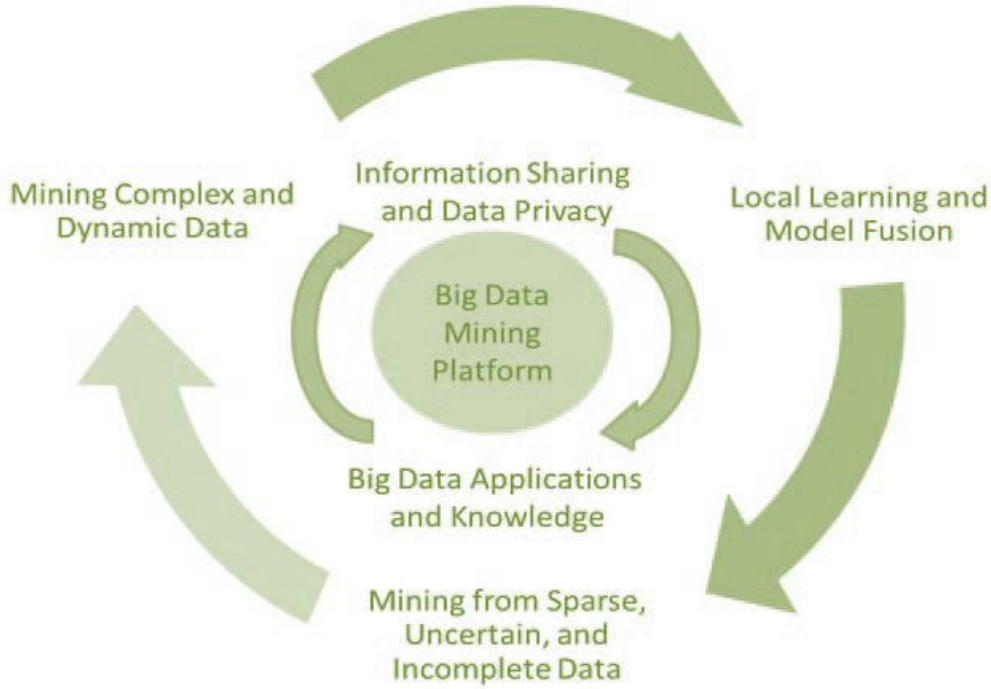

Fig.1 A Big Data Processing frame work

Big Data system, which blends both hardware and software components, is hardly available without key industrial stockholders' support. But, for decades, companies have been making business decisions based on transactional data stored in relational databases. Big Data mining offers opportunities to go beyond traditional relational databases to rely on less structured data like weblogs, social media, e-mail, sensors, and photographs that can be mined for useful information. Major business intelligence companies, such IBM, Oracle, Teradata, and so on, have all featured their own products to help customers acquire and organize these diverse data sources and coordinate with customers' existing data to find new insights and capitalize on hidden relationships.

III. ENTERPRISE AND BUSINESS LOGIC MODEL

As Business requirements process defines the business needs of the application in terms of various factors those are market oriented, product oriented, quality oriented and time saving tends. This tier acts as an interface between easy chair access platform and an efficient algorithm for mining works as management tool for improvement in scalable factors. By comparing various enterprise skill methods apply efficient one for work faster. To integrate with other enterprise applications used within the organization and to deploy across a variety of networks while meeting strict requirements for security and administration management. Enterprise application model is one of the biggest models in software recent trends and development. Now we see it briefly for the implementation of three tier structure. Enterprise

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

application model defines six approaches for scalable and measurable deployment of project that deal with various entities such as user, technology, programmers etc.

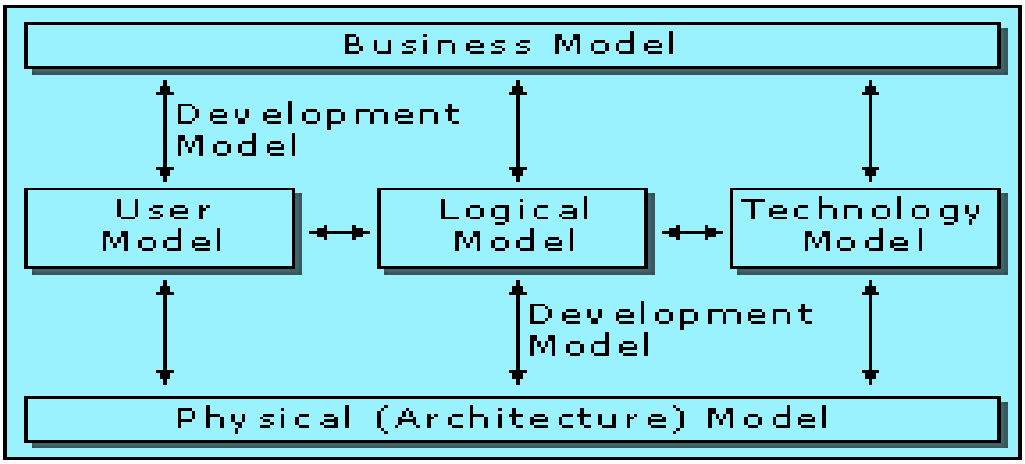

Fig.2 Enterprise Model Approach

I. The Development Model: This model concerned with fast development process for team holders those are working for project. As development process goes on different problem come in front of programmers and solutions for that problems put in such a way that not affects the entire produced system. After solutions got testing milestones are described easily according to the logical and graphical terms with respect to customers and deliverables.
II. The Business Model: The basic goal of organization should be achieved in such a way that extends to time and cost respectively. The requirements of resources needed have been defined as the process of development follows up. Time Constraints followed by all team members because project delivery should not extend. Security measures are defined to provide high reliability and confidentiality to users.
III. The User Model: Now a day's only User friendly systems have a demand in market since system should not be complex and vulnerable to end user. User interface must be in ease-of use in manner. Application support parameter get enhance so that the service oriented feature are in-build in nature. As network connects to system it application may able to run in that and shares information and resources also.
IV. The Logical Model: As the process of development goes on according to period of time so the logical structure of the application should be simple. The application objects and data modelling objects construct effectively and the logic of source code increased.
V. The Technology Model: The development depends on one important factor is technology which may define component development and reuse development tools deployment platforms. The back end database technologies define an efficient technique for storing and managing data in terms of clusters, modules, spread sheets etc. As a project defines in a network, then it should have messaging technologies and effective communication Medias.
VI. The Physical Model: A Physical application architecture distribution and interconnection of components shows the infrastructural issues at low level.

IV. BIG DATA MANAGEMENT WITH MINING

BIG DATA REFERS TO ENORMITY IN FIVE DIMENSIONS:

Volume– from terabytes to peta bytes and up.
Variety– an expanding universe of data types and sources.
Velocity– accelerated data flow in all directions.
Variability– inconsistent data flows with periodic peaks.
Complexity– the need to correlate and share data across entities.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Fig.3 Big Data Characteristic flow diagram

- **Volume:** Many factors contribute to the increase in data volume. Transaction-based data stored through the years. Unstructured data streaming in from social media. Increasing amounts of sensor and machine-to-machine data being collected. In the past, excessive data volume was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use analytics to create value from relevant data.
- **Velocity:** Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations.
- **Variety:** Data today comes in all types of formats. Structured, numeric data in traditional databases. Information created from line-of-business applications. Unstructured text documents, email, video, audio, stock ticker data and financial transactions. Managing, merging and governing different varieties of data are something many organizations still grapple with.
- **Variability:** In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Is something trending in social media? Daily, seasonal and event-triggered peak data loads can be challenging to manage. Even more so with unstructured data involved.
- **Complexity:** Today's data comes from multiple sources. And it is still an undertaking to link, match, cleanse and transform data across systems. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages or your data can quickly spiral out of control.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

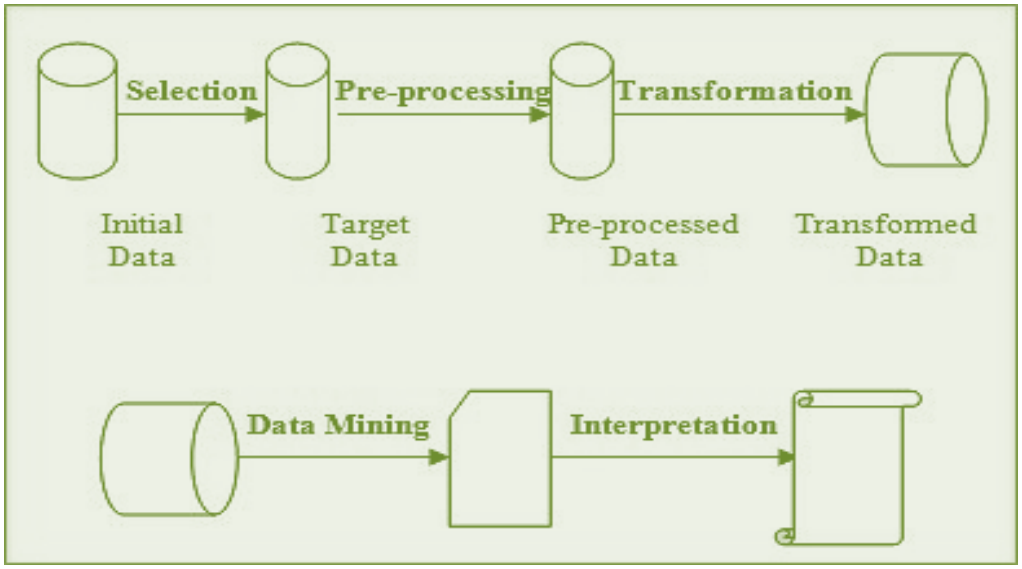

Fig.4 Data Pre-Processing flow diagram

Three key technologies that can help you to get a handle on big data on business logic.

• Information management for big data:
Manage data as a strategic, core asset, with ongoing process control for big data analytics.
• High-performance analytics for big data:
Gain rapid insights from big data and the ability to solve increasingly complex problems using more data.
• Flexible deployment options for big data:
Choose between options for on-premises or hosted, software-as-a-service (SaaS) approaches for big data and big data analytics.

V. CONCLUSION

To explore Big Data, I have analysed several challenges at the data, model, and system levels. To support Big Data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the Big Data. The processing related to big data requires large configurations in terms of storage, processing elements, wide communication media etc. So the processing entities are capable to handle such large and diverse amount of data. As world is survive in a WWW enlargement of data comes from each and every host in a network. But today new emerging technologies make it possible with help of business logic approach to realize the value and importance of big data. By concentrating on the attributes of data we can design a fastest and efficient technique to handle problems related to big data.

REFERENCES

1. Xindong Wu, Gong-Qing Wu, and Wei Ding Data Mining with Big Data IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014.
2. R. Ahmed and G. Karypis, "Algorithms for Mining the Evolution of Conserved Relational States in Dynamic Networks," Knowledge and Information Systems, vol. 33, no. 3, pp. 603-630, Dec. 2012.
3. M.H. Alam, J.W. Ha, and S.K. Lee, "Novel Approaches to Crawling Important Pages Early," Knowledge and Information Systems, vol. 33, no. 3, pp 707-734, Dec. 2012.
4. S. Aral and D. Walker, "Identifying Influential and Susceptible Members of Social Networks," Science, vol. 337, pp. 337-341, 2012.
5. A. Machanavajjhala and J.P. Reiter, "Big Privacy: Protecting Confidentiality in Big Data," ACM Crossroads, vol. 19, no. 1, pp. 20-23, 2012.
6. S. Banerjee and N. Agarwal, "Analyzing Collective Behavior from Blogs Using Swarm Intelligence," Knowledge and Information Systems, vol. 33, no. 3, pp. 523-547, Dec. 2012.
7. E. Birney, "The Making of ENCODE: Lessons for Big-Data Projects," Nature, vol. 489, pp. 49-51, 2012.
8. J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," J. Computational Science, vol. 2, no. 1, pp. 1-8, 2011.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

9. D. Howe et al., "Big Data: The Future of Biocuration," Nature, vol. 455, pp. 47-50, Sept. 2008.
10. B. Huberman, "Sociology of Science: Big Data Deserve a Bigger Audience," Nature, vol. 482, p. 308, 2012.
11. "IBM What Is Big Data: Bring Big Data to the Enterprise," http:// www-01.ibm.com/software/data/bigdata/, IBM, 2012.
12. A. Jacobs, "The Pathologies of Big Data," Comm. ACM, vol. 52, no. 8, pp. 36-44, 2009.
13. I. Kopanas, N. Avouris, and S. Daskalaki, "The Role of Domain Knowledge in a Large Scale Data Mining Project," Proc. Second Hellenic Conf. AI: Methods and Applications of Artificial Intelligence, I.P. Vlahavas, C.D. Spyropoulos, eds., pp. 288-299, 2002.
14. A. Labrinidis and H. Jagadish, "Challenges and Opportunities with Big Data," Proc. VLDB Endowment, vol. 5, no. 12, 2032-2033, 2012.
15. Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," J. Cryptology, vol. 15, no. 3, pp. 177-206, 2002.
16. W. Liu and T. Wang, "Online Active Multi-Field Learning for Efficient Email Spam Filtering," Knowledge and Information Systems, vol. 33, no. 1, pp. 117-136, Oct. 2012.
17. J. Lorch, B. Parno, J. Mickens, M. Raykova, and J. Schiffman, "Shoroud: Ensuring Private Access to Large-Scale Data in the Data Center," Proc. 11th USENIX Conf. File and Storage Technologies (FAST '13), 2013.
18. D. Luo, C. Ding, and H. Huang, "Parallelization with Multiplicative Algorithms for Big Data Mining," Proc. IEEE 12th Int'l Conf. Data Mining, pp. 489-498, 2012.
19. T. Mitchell, "Mining our Reality," Science, vol. 326, pp. 1644-1645, 2009. [37] Nature Editorial, "Community Cleverness Required," Nature, vol. 455, no. 7209, p. 1, Sept. 2008.
20. J. Shafer, R. Agrawal, and M. Mehta, "SPRINT: A Scalable Parallel Classifier for Data Mining," Proc. 22nd VLDB Conf., 1996.
21. K. Su, H. Huang, X. Wu, and S. Zhang, "A Logical Framework for Identifying Quality Knowledge from Different Data Sources," Decision Support Systems, vol. 42, no. 3, pp. 1673-1683, 2006.