Big Data Issues and Challenges

S. Naveenkumar
(Ph.D) Category –B
Dept. of computer science, (R&D),
Bharathiar University, Coimbatore
Tamilnadu, India. Pin – 641 046

Dr.S.Vijayaragavan ,Ph.D.
Professor and Head,Department of CSE,
Paavai Engineering College,
NH-7, Pachal, Namakkal, Tamilnadu, India, Pin- 637018.

Abstract - The aim of this topic is to identify and classify the main issues and challenges facing different business organizations when implementing big data technologies. This is highly significant given the theoretical and practical implications and importance of such technologies. It is also important to highlight these challenges given the scarcity of research in this domain despite its value. To determine big data issues and challenges, a qualitative research methodology was followed. The big data problem gathered from data scientist and other top level managers were analyzed by utilizing a bottom-up content analysis technique where content is coded and classified so as to let concepts emerge naturally. The study results show that big data issues and challenges are both managerial/organizational and technological. According to the results of the study,

The main managerial/organizational big data challenges are:

- **Top level commitment and support**
- **Project supporter**
- **User input**
- **Team ability and work**

The technological challenges are:

- **The selection of big data architecture**
- **The creation of the enterprise schema for big data**
- **Data integration and scalability**
- **Data quality**
- **Security and privacy risks**
- **Networks and telecommunications**

Keywords – Big data issues

I. INTRODUCTION

The digital era has intended that the accessibility of appropriate information and knowledge have become critical to the success of the industry. The next information uprising is about information content and its purpose .However, organizations need to adapt in order to endure and succeed as their business domains, processes and technologies change in a world of increasing environmental complexity. Enhancing their performance and competitive positions by improving their ability to respond quickly to rapid environmental changes with high quality business decisions can be supported by exploiting technologies such as big data and its analytical tools Nevertheless, assuring business intelligence basically requires having relevant, reliable, accessible, accurate, timely, complete, coherent, and consistent quality information to the decision at hand . Hence, business intelligence through decision making improvement is a major concern for business managers nowadays. No doubt, accessing large and vast amount of the data stored in business organizations operational systems has become progressively more time consuming and cumbersome .Hence, business organizations have embarked on data warehousing to overcome these problems through integrating heterogeneous operational data sources. Big data provides a technological infrastructure enabling business organizations to extract data from source systems, store the data, and get the respective/quality of data from the stored data for analytical purpose .Big data is considered one of the most powerful decision support

and business intelligence technologies that have emerged in the recent years. Nevertheless, the realization of big data benefits by business organizations has been below opportunity. Hence, this study is mainly focused on two points: first, it presents, illustrates, and discusses the role and value of big data as an aspect or driver for decision support system, and secondly it is critically analyzing both organizational and technological issues and challenges of big data development/implementation with current approaches and technologies. The remainder sections of this study are well thought-out as follows. In the next section, we provide the research goals and objectives. Then, a theoretical background about the study area is provided by illustrating the main area concepts followed by discussing the main motives behind deploying big data. Thereafter, the research methods are discussed. Then, the main challenges and issues of big data development are discussed as the main results of the study. Finally, the study conclusions are presented.

II. RESEARCH GOALS AND OBJECTIVES

Big data technologies and projects are highly significant to business organizations in terms of investment, time, and effort in addition to their perceived value. However, the failure rate of such projects is high. We therefore aim – in this study- to identify the main issues and challenges facing different business organizations throughout the implementation of big data technologies. Achieving this goal is significant as the results can be used as a roadmap or guidelines for software organizations aiming to implement big data technologies. The objectives of this research are multifold. First, we aim to determine various decision-making and organizational factors those deemed significant to the success or failure of big data implementations. It is very important to highlight these factors as many decision makers within organizations believe that big data projects are purely technical ones and thus decision-making and organizational aspects and domains are irrelevant. Given the technical complexity of such technologies and projects, there are also some technological challenges to be determined and this represents our second objective. Thirdly, we aim to provide a rock-solid and clear academic background about big data, business intelligence and decision support system concepts highlighting the major drivers for implementing big data technologies and its relationship with business report represents our fourth objective in this research.

III. THEORETICAL BACKGROUND

A. Introduction to Big data

Nowadays one of the key developments in the Information System field is big data. Unlike On-Line Transaction Processing (OLTP) databases, that is application-oriented, detailed, and operational. Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and information. The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set. Big data is a large volume unstructured data which cannot be handled by standard database management systems like DBMS, RDBMS or ORDBMS .Big data is "a volume, variety, velocity, variability, veracity, complexity and non-updatable collection of data to support management decision-making processes and business intelligence." Big data are widely perceived as valuable devices for acquiring information from multiple sources and delivering it to managers and analysts who may be using different software tools or computer platforms with special features and capabilities. Big data meant to support managers with answers to important business questions that require analytics such as pivoting, drill-downs, roll-ups, aggregations and data slicing and dicing. Moreover, all levels of management decision-making processes are supported by big data through the collection; integration, transformation, and interpretation of internal, external and other type's data Moreover have elaborated how a big data could provide useful and valuable information and knowledge at a strategic, management control, Knowledge and operational levels.

Table 1:

Volume	The quantity of data that is generated is very important in this context. It is the size of the data which determines the value and potential of the data under consideration and whether it can actually be considered Big Data or not
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 the 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 analyze 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
Veracity	The quality of the data being captured can vary greatly. Accuracy of analysis depends on the veracity of the source data.
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 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

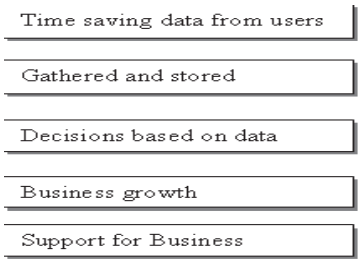
Before digging more into the main study objectives, it's crucial to differentiate a big data, the repository of collective data from big data which revolve around the development, management, methods, and a practice that defines how these summarized data are acquired, integrated, interpreted, managed, and used within business organizations. On the other hand, business intelligence is deep-rooted in interpreting the data acquired through environmental scanning with respect to a business task "contextualization" and it's supposed to provide tactical and strategic information to decision-makers so as to be able to manage and coordinate operations and processes in their business organizations. For the purpose of business intelligence, many analytical tools have been developed such as: excel, reporting tools, dashboards, OLAP, and data mining. Business intelligence revolves around knowledge discovery and inferences by analyzing the data stored in big data to acquire valuable information; in other words, "Big data is a repository of intelligence from which decision making can be derived". Another important issue is the differentiation between Operational Data Store (ODS) and big data. Although ODS uses big data technology to provide integrated view of data, but its intended to assist day-to-day operations and not decision making, since ODS is a volume, variety, velocity, variability, veracity, complexity and volatile (updatable) data store that contain only business organization detailed data for operational usage. Furthermore, the differentiation between the data stored in a big data and the metadata is important to both big data users and maintainers. Metadata (data about data) provides the following information about the big data:

I. A directory about the data stored in a big data (location, index, etc.).
II. A guide to mapping data from operational sources to big data form.
III. Summarization rules.

B. Reasons Underlying the Implementation of Big data Technologies

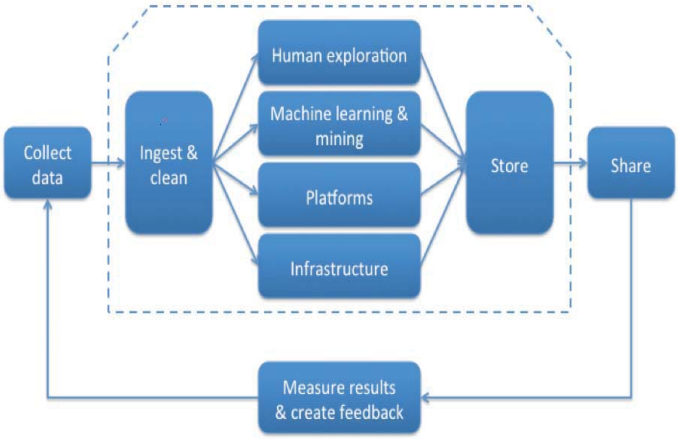
A wide variety of tangible and intangible benefits can be gained from big data applications "see figure 1" big data is a data repository which is relevant to the management of business organization and from which the needed information and knowledge to effectively manage the organization are emerged Initially, big data was viewed as a way by which business organization could solve the problems associated to their independently legacy systems which often contains inaccurate, duplicate, and dissimilar data about the same entity. Big data technology (figure 2) can help managers make more effective decisions by providing them with suitable information which is fundamentally different from the type of information that software industry use in their day-to-day information.

Figure 1:
The Benefits from big data



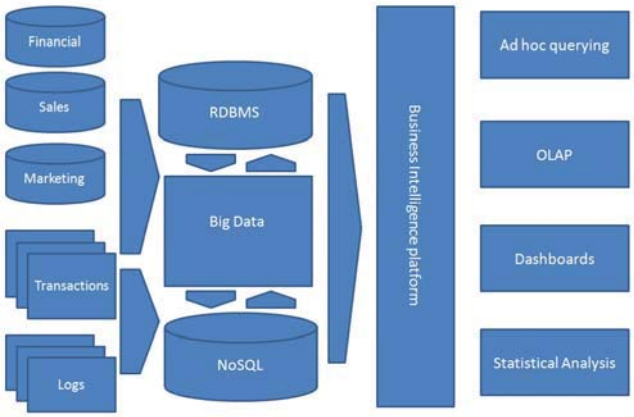
Big data contain cleaned, aggregated, consolidated large volumes of data that is accumulated in structured, semi-structured and unstructured data structure to support the decision support system analysis. Not only a big data recognizes the need for current and future data, but also recognizes the need for historical data for instance, trend analysis requires a great deal of historical data regardless the company. Big data allows a business organization to manipulate a great deal of data in ways that are useful to it, such as: cleansing, organizing, describing, summarizing and storing large volumes of data to be altered, analyzed and reported .

Figure 2:
Big data supply chain



Big data offers effective service data management and data delivery processes by expanding to the knowledge into cross-functional integrative decision support system. Accordingly business organizations could compete better by having the ability to learn from the past, to analyze current situations, and to predict the future scenarios.

Figure 3:
Generic big data architecture



Business organizations could compete better by having the ability to learn from the past, to analyze current situations, and to predict the future decision making analysis. The big data provides the large scale IT infrastructure for contemporary decision support and business intelligence (see figure 3). They argue that the main reasons behind that is the use of structured, semi-structured and unstructured" to organizes large data sets in

ways that are meaningful to managers besides being relatively easy to query and analyze. It has been proved that MDDM is the most suitable for On-Line Analytical Processing "OLAP" applications, data mining and advanced reporting functions. The conceptualized MDDM can be physically realized in two ways: first, by using trusted relational databases "star schema/snowflake schema" or, by using a specialized multidimensional databases. Many computer assisted analytical processes such as: data mining and OLAP that used to analyze data from different angles and distilling it into actionable information run over big data. Moreover, the category of big data project is IT infrastructure type that provides a foundation for IS/IT application development (i.e. ERP, CRM) which could provide business organizations with strategic competitive advantages. Another potential benefit of big data is that using a single data source may facilitate business process re-engineering at business organizations. The big data is the way in which business organization converts its data into information which can be represented into different ways (textual, graphically, etc.) based on its reporting capabilities. Also, they argued that ad hoc system which is provided to managers based on big data allow them to generate speculative information such as projections, and allow them to explore "what-if" analysis. Thus, the desire to improve decision-making and business performance has been the fundamental business driver behind big data. Furthermore, big data analytical processing and its key role is to offer forceful business intelligence to business organizations' decision-makers, through enriching their abilities in understanding business problems, exploiting opportunities and improving business performance through big data. Hence, by leveraging big data technology for business intelligence initiatives, business organization can gain strategic competitive advantage.

C. The Role of big data in Enhancing Business Performance

Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The 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.

IV. RESEARCH METHODS

Despite the importance of big data technologies in theory and practice, only few research have been conducted to study the big data challenges and success factors. Therefore, in this research, we follow a qualitative paradigm so as to comprehensively and clearly identify the main issues and challenges facing business organizations when implementing big data technologies. A semi-structured interview protocol was used with big data project managers and senior members with both technical and managerial backgrounds and positions in order to collect relevant data. The duration of discussion was about 4 hours in every week end for four months. The main themes discussed within the interviews were "the role of management in data big data implementations", "the role of team members in big data implementation", "the role of users in big data implementations", "the complexity of big data technologies", and "technical and organizational requirements of big data". The collected data were analyzed following a content analysis technique. Content analysis can be defined as any research technique for making inferences by systematically and objectively identifying shared common properties for the phenomenon under investigation. To end up with systematic and objective inferences, a data classification technique where data are read and categorized into concepts that are suggested by the data rather than imposed from outside is mandatory .To be aligned with the aforementioned principles, we analyzed the collected data the systematically .In fact, a bottom-up open coding procedure was followed here where interviews have been broken into segments or incidents which are found in a phrase . Each incident was coded by an indicator characterized as short and effective. The indicators of different incidents were compared to each other to develop indicators of a higher level (i.e. indexes). Thereafter, the codes or indexes were analyzed where those pointing towards the same theme were aggregated together .Aggregating the codes based on themes allowed big data concepts to emerge. Then, these concepts were categorized into two different classes: technological challenges including concepts and managerial/organizational challenges including concepts (see Table 2).

TABLE 2: ISSUES AND CHALLENGES OF BIG DATA IMPLEMENTATIONS

Managerial/Organizational Challenges	Technological Challenges
• Management promise and carry • Project supporter • User Involvement and Participation • Team Skills and Composition	•The Selection of big data Architecture • The Creation of the Enterprise Schema • Data Integrity and Scalability • Data Quality • Human-Computer Interfaces • Mining the big data • Security and Privacy Risks •Networks &Telecommunications

V. RESULTS AND DISCUSSION: ISSUES AND CHALLENGES OF BIG DATA DEVELOPMENTS AND IMPLEMENTATIONS

Although many business organizations appear to have implemented big data, the road to its success has been plagued with failures. In addition to the high cost and resource requirements needed for such project, many other societal/cultural, organizational and technical reasons may be responsible for their high failure rates. Implementing large-scale IS/IT projects such as big data is complex undertaking; a variety of many organizational and technological challenges and issues facing business organizations when they starting big data projects. This could be referred to the fact that big data project is an enterprise-wide scale initiative that involves acquisition and/or development of tools and applications for user access, database maintenance, and data transfer and scrubbing. On the other hand, lack of organizational readiness, project size and structure, incomplete risk management, familiarity with technology, and skills of system analysts can negatively impact the project's outcome. In the following sub-sections, both managerial/organizational and technological issues and challenges have been discussed in detail.

A. Managerial/Organizational Issues and Challenges

a. Management promise & carry

The managerial commitment through senior management's participation, involvement, and support is a key element for an innovation such as big data to be adopted and subsequently used efficiently and effectively. Organizational commitment ensures the consistency between big data project and strategic vision and direction as well as shows the importance of this innovation to business value in order to sustain favorable. Moreover, if big data is backed by the organization's organization, the users are more likely to accept. On the other hand and since big data is a high multiyear investment, significant politically charged issues could be triggered; thus the senior management and other stakeholders' attention, involvement, support, and commitment is highly need.

b. Project supporter

The planning process for big data initiative falls very much into a broader strategic planning exercise for municipal Information Systems. Therefore, big data project needs proper executive sponsorship such as a champion at the top management level to fund the project toward its completion and to be properly maintained. When a champion is strongly supportive, users are willing to acquire knowledge about big data technology continuously even after the introductory training. Thus, champions exhibit "transformational leadership behavior"; in other words, when they support the big data project, they possess the needed skills to overcome resistance that may arise within the organization while implementing the big data system.

c. User Involvement & Participation

Generally speaking and in order to get the most from big data, users must endure continual, formal, and systematic training for such large-scale initiative; this will enable users to understand functions supported by big

data much better as well as being more accountable for making big data produce higher quality information. User involvement is crucial especially when the requirements for a system are unclear. Moreover, such training will be useful in eliminating or at least reducing users' resistance. If users are not willing to use big data, then it's necessary to change their attitudes by offering them a suitable motivation and/or giving them additional training. The assessing the talents and mind-sets of big data users is a necessary preliminary step to determine the kind of training them need and their readiness for such computing practice. Hence, if the big data project to go forward, both management and users must be convinced of its value and its deliverables.

d. Team Skills & Composition

The big data development team skills have a major impact on its outcomes. The big data project already had a great deal of top management support, users and IS staff should be part of the big data project team and should be viewed as partners with common objective of creating a useful and reliable big data. They also argued that when users are part of the team, they will be more proactive and patient in helping to exploit such a technology as well as understanding its capabilities and limitation and being more able to provide valuable inputs into future enhancements. Thus, the adoption of big data could be facilitated by ensuring the existence of key users and stakeholders within the project team since they can creatively identify ways through which needed knowledge can be extracted from multiple functional areas within the business organization. In addition of getting a good combination of members in the big data project team, getting the right person to lead the big data project, who is technically competent, has adequate business knowledge and interpersonal skills is crucial to big data success .

B. Technological Issues and Challenges

Obviously, a big data development with its supporting middleware and analytical tools (i.e. OLAP, data mining, and dashboards) represents a large investment for business organization. On average, a big data installation cost is $ 1 million and in some cases exceeds $10 million. Thus, a great deal of care must be taken to determine the nature of its ultimate development and use. In this section, a number of technological issues and challenges in building a big data are discussed.

a. The Selection of big data Architecture

There are two fundamental approaches to big data: enterprise-level big data (top-down approach) and departmental-level data marts (bottom-up approach). Unlike data marts which fulfill the information needs of a discrete group or a specific department, big data are designed to satisfy the information needs of an entire big business organization thus being more complicated and expensive to build and to maintain. Data marts main advantages are: it requires less time to be built, thus delivering value to customer more quickly. Further, it reduces the risk associated with such initiative since its design and development is simpler, and the amount of investment is lower than in big data. On the other hand, data marts within the same enterprise are often different, thus integrating them in an enterprise-wide big data can be a cumbersome and challenging process.

b. The Creation of the Enterprise Schema

The step that follows the selection of the big data architecture is its logical design. Since big data represents an integration of multiple databases exist within a business organization, their existing schemas could be used as a foundation from which the new enterprise data model will be generated and designed. These existing schemas will be valuable in determining the main entities and their relationships although designers need to refine them and to add any missing elements. The following challenges and issues regarding the enterprise schema creation have been identified as follows.

1. Structural heterogeneity: given that the enterprise data model relies heavily on the overlapping existing different database models, the structural heterogeneity arise. To give only one example, the same entity which represents the same real-world class may exist in multiple database schemas with the same fields, but the fields' size or data type may differs.

2. Semantic heterogeneity: Semantic heterogeneity exists when data is defined differently indifferent source schemas, thus pose enormous challenges. This includes synonym and homonym problems. The first represents

having two different names in different database schemas for the same real-world class, and the later occurs when having the same name for different real-world classes in different database schemas.

3. Constraints mismatches: generally speaking, there is no right approach for resolving mismatch incompatibility; such as having manager. Salary >= 10000 in one database, and having manager. Salary>=15000 in another database.

c. Data Integration & Scalability

One of the significant technological challenges to big data development is data integration from heterogeneous sources. The wide variety of critical information requirements for business decision-making and their different sources; both internal {spreadsheets, documents, plain files, etc.} and external {governmental documents, Journals, etc it represents a great challenge to big data designers. However, the internal information is useful for tactical level decision making and evaluation, whilst the external one is useful for strategic decision making and evaluation. The big data incorporation of unstructured and semi-structured information about partners, policies, rules, competitors, etc. through an environmental scanning process influencing the big data usefulness for decision making purposes. On the other hand, the integration of past, present and future information is also challenging. Most often, data has to be extracted from a wide variety of distributed operational information systems that operate on different hardware platform and use different Database Management Systems (DBMS) with different structure. Thus, data integration through the need to handle such heterogeneous sources of data leads to considerable complexity. To give just two examples, determining whether a pair of records coming from two different databases represent the same real-world entity or not is a great challenge, another example is when two records coming from two databases having the same entity identification (primary key), but the other fields' values are different .Hence, big data function represented the most expensive and time-consuming portion in the big data development. Generally speaking, the common big data development method which relies on capturing and acquiring a large volume of transactional data and loading it into a big data is not sufficient since not all information would be stored in the big data, not all departments would desire the same information with the same physical format, and not all information would be suitable for decision-making purposes. Hence, data may remain relatively useless for business intelligence and decision-making purposes even when it is efficiently acquired and stored. For these reasons, performing information system requirements analysis could be useful in determining the types of information required for effective management of business operations, control, strategic planning, and decision-making. Furthermore, one of the significant challenges for the big data research community nowadays is that big data must be understandable and adaptable, and must include experience-based organizational knowledge in order to provide information that improves business managers' ability in understanding and identifying situations requiring actions as well as enabling them to understand decisions' implications and their impacts over time. Moreover, big data stores non-volatile information over time, thus scalability issue should be taken into consideration. Big data must be designed for change from the beginning since it has a periodic updates rate of few terabyte (TB).

d. Data Quality

Another major concern for big data is data quality. Further, it has a rigorous impact on the overall business organization's performance. Thus, relying on inaccurate, incomplete, inconsistent, imprecise, irrelevant, and non-coherent information for decision making purposes in general is a disaster, and it is more harmful for strategic level of decisions. Loss of information, insufficient information (ambiguity), meaningless data, and incorrect data have been identified by as the most observed data problems. Interestingly, they also indicate that there is a proportional relationship between the design & production techniques involved in generating the data and the quality of resulted data. To give just one example, approaches for cleansing data during big data have been shown to be successful.

TABLE 2:NOTABLE DATA QUALITY MEASUREMENT

Measurement	Cited	Measurement	Cited
Accuracy	25	Content	2
Format	4	Freedom from bias	3
Comparability	2	Relevance	16
Reliability	22	Efficiency	3
Interoperability	4	Informativeness	2
Conciseness	2	Completeness	15

Timeliness	19	Precision	2
Importance	12	Clarity	12
Level of detail	24	Understandability	2
Sufficiency	3	Consistency	5
Quantitativeness	45	Scope	72
Usableness	5	Flexibility	8
Usefulness	2	Currency	2

Although disagreement in the literature on data quality dimensions (see table 2), assuring better quality requires thorough understanding of its meaning and dimensions. However, more recent researchers have stated accuracy, timeliness, completeness, and consistency as quality dimensions. Nevertheless, the concept of data quality is relative and the quality of data is of high subjective which makes data quality assurance more challenging. Thus, more attention to data quality issues in big data is needed.

e. Human-Computer Interfaces

One of the paramount important issues and the primary determinant of big data success from the end-user perspective is the human-computer interface. Providing more user-friendly interface tools such as pull-down menus, menu-based interfaces and drags and drop capabilities for analysis and reporting purposes consider one of the main competitive advantages for big data vendors. The major reasons for not using big data is that users perceived it as more technically-oriented reporting tools, thus depending on 'power-users' to get an answer to a query. However big data supports ad-hoc quires, thus big data designers should trade-off between the ability to express such queries and its ease-of-use. Finally, for big data to be successful and to enhance business intelligence mainly by supporting organizational decision-making, users must be shielded from its underlying complexity by a user-friendly human-computer interface.

f. Mining the big data

In order to perform strategic analysis of big data, a number of tools have been developed such as: data mining, OLAP, trend analysis, forecasting and simulation. These tools are useful to extract a valuable hidden knowledge from large amount of data through slicing, dicing, rolling-up, and drilling down capabilities. To give just one example, data mining, or Knowledge Discovery from Databases (KDD), techniques could be applied in Customer Relationship Management (CRM) to predict customer profitability, to conduct click stream analysis, and for customer segmentation purposes. However once a problem is identified, the available commercial tools are not very effective for generating solution alternatives nor in discovering useful knowledge for strategy formulation and implementation. On the other hand, these tools are still considered complex and somehow sophisticated to use from the end-user perspective.

g. Big data Security & Privacy Risks

Most often, all enterprise data will be stored in its big data. Data in a big data is orderly, integrated, centrally-located and easily accessible thus offers an appealing target .However, the data in a big data shouldn't be threatened, lost, and/or manipulated. Moreover, the information and the knowledge extracted from a big data should be reliable; otherwise it can have disastrous effects on the enterprise. However, the quality of knowledge extracted from a big data is totally dependent on the quality of data stored in it. Hence, many techniques such as back-up, disaster recovery plans, strong password policy, Intrusion Detection Systems (IDSs), firewalls, encryption and anti-virus SW that reduce the risks of losing or manipulating the enterprise data should be implemented to mitigate the risks of losing this valuable asset. Furthermore, end-users' awareness of these risks and security procedures is highly beneficial. On the other hand, current privacy approaches such generalization, condensation, randomization, cloaking, etc. are all special cases to protect data by folding actual data values into patterns which in fact comes at the cost of some imprecision. Thus, more general significant approach is required.

h. Networks & Telecommunications

Since a big data serves most, if not all, departments in an enterprise, the ability to access it efficiently and effectively is a major concern. Sometimes, big data users are located in different regional areas and proper network connectivity should be available to allow them to access the big data effectively and reliably. Moreover, network bandwidth and its selection criteria is major concern when big data users are distributed among different sites.

VI. CONCLUSIONS

Creating a big data has itself proved to be difficult and problematic and it is highly perceived as high-risk/high-return initiatives. The researcher analysis reveals that despite agreement in the information system literature on the importance of big data to an organization success through enhancing its decision-making quality, the attainment of such business intelligence based on it is still poorly proved empirically. Moreover and although the wide variety of motivators mentioned in the IS literature for developing big data, overcoming the problems in the legacy heterogeneous systems as well as obeying to the governmental regulations are the most two conceivable reasons for which organizations are building their big data. The big data is highly recognized as an infrastructure; many applications can run over it such as CRM and DSS systems. On the other hand, many techniques, such as data mining, OLAP and dashboards have been rising to prominence to extract business intelligence from big data. Furthermore, big data meant to be used by managers since they support decision-making process. Nevertheless, these techniques are still not very effective and are highly perceived as technically oriented by the end-users. Nevertheless, big data has experienced relatively high failure rates and its spread and/or use has been to some extent limited. Perhaps due to the facts that designing and developing a big data is a risky, costly and complex process. It requires a huge amount of money as an investment, spans over years, and needs a wide variety of technical and managerial skills. Generally speaking, social aspects are shaping the technology. Hence, the interaction of technology and social context is the key determinant of big data. Nevertheless, despite the technical complexity of big data design and implementation, social/cultural and organizational factors are the most cited reasons behind big data failures. From these insights and conclusions, each and every societal/cultural, organizational and/or technological issue and challenge mentioned previously represents a valuable area for research purposes. Interestingly, it is recommended to focus on areas such as improving business intelligence techniques in terms of user-friendliness and effectiveness in further researches. Other interesting and challenging area for research purposes is the integration of structured, semi-structured and even unstructured data in big data.

VII. ACKNOWLEDGMENT

I am thankful to my guide Mr.Dr.S.Vijayaragavan (Professor, Centre Head, Department of CSE), Paavai Engineering College, Namakkal, Tamilnadu, India for his valuable guidance ,encouragement and co-operation during the course of this conference and its presentation.

I am also thankful to my parent, wife and god for helping opportunity to me to present this conference.

REFERENCES

[1] https://www.safaribooksonline.com/library
[2] https://www.academia.edu
[3] http://en.wikipedia.org/
[4] Drowning in numbers -- Digital data will flood the planet—and help us understand it better. The Economist, Nov 18, 2011. http://www.economist.com/blogs/dailychart/2011/11/big-data-0
[5] Using Data for Systemic Financial Risk Management. Mark Flood, H V Jagadish, Albert Kyle, Frank Olken, and Louiqa Raschid. Proc. Fifth Biennial Conf. Innovative Data Systems Research, Jan. 2011.
[6] Pattern-Based Strategy: Getting Value from Big Data. Gartner Group press release. July 2011. Available at http://www.gartner.com/it/page.jsp?id=1731916
[7] Gantz, J. and E. Reinsel. 2011. “Extracting Value from Chaos”, IDC’s Digital Universe Study, sponsored by EMC.
[8] Hsinchun Chen , Roger H. L , Veda C., “BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT”, MIS Quarterly Vol. 36 No. 4/December 2012,Eller College of Management, University of Arizona, Tucson, AZ 85721 U.S.A.(2012).
[9] Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram, W. , “Shared disk big data analytics with Apache Hadoop” , Dec.,2012.
[10] Stephen Kaisler,Frank Armour Big Data: Issues and Challenges Moving Forward, 46th Hawaii International Conference on System Sciences, 2013.
[11] The Economist. 2010. “Data, Data Everywhere”,(online edition, February 28) http://www.economist.com/node/15557443