INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Analysis and Optimization: The Real Stuff of Business Competitive Advantage from Cloud Big Data Analytics

Gopikrishna Tatapudi
MIS Department, College of Business Administration, King Saud University, Riyadh, Saudi Arabia
Mohammed AlSudairi
MIS Department, College of Business Administration, King Saud University, Riyadh, Saudi Arabia

Abstract:
The success of enterprises in the 21st century in an increasingly competitive market place depends critically on the quality of knowledge which those enterprises apply to their key business processes. The challenge of deploying the knowledge assets of an enterprise in creating competitive advantage and ensuring a sustained business edge is important. Despite the importance to researchers, managers and policy makers of how information technology (IT) contributes to organisation performance and generates business value based on the resource-based view of the firm that integrates various standards of research into a single theoretical framework, in this article, an attempt is made to describe the relationship between the competitive advantage through theoretical strategic views and cloud based big data analysis & optimization platforms. So the objective of this paper is to provide a comprehensive understanding on the cloud big data analytics scenario with understanding on some of the buzz words. It provides a comprehensive approach to derive the patterns and analysis to support business strategy towards enterprise management and decision making by enabling cost reduction strategies at storage level costs as well as processing and analysis costs. It is a journey of effort of to compare the results of a survey on human executive opinion results versus by big data analytics results. For this purpose, a questionnaire is prepared, which is sourced from the one designed by Deloitte's 2012 global outsourcing and in sourcing survey.

Keywords*: Achieving Business Competitive Advantage; Big Data Analytics; Business Intelligence; Cloud Computing; Exploring Innovative Business Strategy; Enterprise System Management*

1. Introduction

According to Valencia & Clemmons (2011-2012), achieving clarification on the business strategy becomes important before planning an enterprise. Further, Information flow becomes one of the most crucial problems while planning for enterprise strategy. Thus it might ask for inducing change in organisational structures not just computer systems to address information-flow problems in a way to provide competitive advantage with proper enterprise information flow (Davenport, 1998). The success of enterprises in the 21st century in an increasingly competitive market place depends critically on the quality of knowledge which those enterprises apply to their key business processes. The challenge of deploying the knowledge assets of an enterprise in creating competitive advantage and ensuring a sustained business edge is important. The potential values that can be generated out of knowledge management which can act as drivers of enterprise business are: competition, customer focus, managing mobility of work force, understanding, developing equity in the workplace and the global imperative. Thus the role of knowledge as the key source for competitive advantage in enterprise has become a hot debated topic (Ndlela & du Toit, 2001).

2. Problem Statement and Objective

Despite the importance to researchers, managers and policy makers of how information technology (IT) contributes to organisation performance and generates business value based on the resource-based view of the firm that integrates various standards of research into a single theoretical framework, in this article, an attempt is made to describe the relationship between the competitive advantage through theoretical strategic views and cloud based big data analysis & optimization. So the objective of this paper is to provide a comprehensive understanding on the cloud big data analytics (CBDA) scenario with understanding on some of the buzz words. CBDA provides a comprehensive approach to derive the patterns and analysis to support business strategy towards enterprise management by

enabling cost reduction strategies at storage level costs as well as processing and analysis costs. It is a partial effort of comparing the results through a survey on human executive opinion results vs. by big data analytics results.

3. Literature Review

Earlier research literature works that has attempted to work in similar kind of problem are mentioned as follows: while the work of Melville, Kraemer & Gurbaxani (2004) research provides the macro view, the work of Powell & Dent-Micallef (1997) investigates linkages between information technology and firm performance. When e-business is understood as the contemporary style and trend of doing business, it has been recognized during the 21st century enterprise management, AlSudairi & Vasista (2012) developed an integrated theoretical framework that essentially provides a comprehensive idea of considering both tangible and intangible resources and internal and external views of the enterprise. However it argues that a service based view in association with resource based view will give an opportunity to enterprise to better deal the system dynamics with a concept of what is called transaction-cost economics (TCE). Because it is believed that a firm's organisation structure should 'fit' it's strategic position, which characterized by the customer needs it serves and the technology and asset portfolio it employs, in that it adopts a transaction cost economizing form of organisation given its strategic position (Nickerson & Silverman, 1997). The transaction cost insights can become useful complements to existing approach to strategic management. Transaction costs are essential aspects of processes of creating, capturing and protecting value. If the transactions costs are zero, these processes do not pose any strategic problems. Transaction costs become positive when inefficiencies causing by transaction costs are existing in the enterprise (Foss, 2003). Cloud computing has a potential to implement the concept of managing transaction cost economics of the firm by giving an opportunity to provide an unified view on business insight on the recorded transactions. For example, Knowledge on Transaction cost theory essentially helps organisations to consider strategic decisions on whether to go for outsourcing or to assign and make use of internal resources contextually.

Cloud computing is an emerging new computing paradigm for developing computing services. The approach relies on a number of existing technologies e.g., internet, virtualization and grid computing. It follows pay-as-you-go (an implied follow of transaction cost economics approach) computing service approach and makes small and medium enterprises commercially viable in the current climate of economic difficulties (Sultan, 2011).

Importance of the IS/IT Outsourcing both in public and private sector can be considered as a multi-disciplinary study nature. AlSudairi & Dwivedi (2010) have made an effort to conduct a systematic survey of the literature pertaining to research on IS/IT outsourcing.

The logic of governance and public policy design are macro level issues and follow enterprise management approach.

Enterprise management mostly relies on unbounded/open rationality. However firm level performance as an object relies on bounded rationality. Further the relationship between micro level issues based on bounded rationality and macro level issues based on openness can be well understood (Anonymous, Undated, p. 6). Hence in order to better deal the enterprise management what is needed is an open and flexible strategic development strategy that can manage the bounded rationality based firm level performances in order to better address and manage the context based scenarios. This is what can be called as the 'essence of enterprise management theoretical framework'. It is believed that Cloud computing has that potential to provide such flexibility and versatility of managing enterprises in the scale range of small and medium enterprises to large governments even under collaboration strategy serving with unit of analysis capabilities (Dyer & Singh, 1998).

4. Methodology

Organisations need to consider the value of associating the quality consciousness and service levels in addition to the cost and duration of delivered services. Decision Support Systems in Cloud enables scale, scope and speed economies. Use of agile based service-oriented decision support systems (DSS in cloud) is one of the major trends for many organizations. For many companies especially to small and medium size enterprises, the pay-as-you-go service-oriented computing model (which follow cloud computing model) has been becoming very attractive (Demirkan & Delen, 2013). Contemporary businesses are adopting Cloud computing where consumption of IT as a service is enabled.

Coupling the cloud with big data phenomenon, organisations will get opportunities to make a decision regarding consuming IT as an external service versus internal infrastructure investments.

IDC defines 'Big data technologies' as a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery and/or analysis (Gantz & Reinsel, 2011). Push based Big Data Analytics adopts Extract-Transform-Load processing to automatically analyze un-modelled data in order to consume data of interest more rapidly. Stream processing software supports real time analytical applications designed to continuously optimize business operations. Automated analysis and filtering is needed for rapid consumption of data (Ferguson, 2012). Big data is a horizontal cross-section of the digital universe and can include transactional data, warehouse data, metadata and other data residing in huge files from domains such as media/entertainment, healthcare and video surveillance etc as well as from social media (i.e. from Facebook, Twitter etc.) as data sources. Big data will inject high-velocity requirements associated with capture and analysis to generate results or predictive reporting so as to generate and show business opportunities. Cloud based big data platforms will make the small and medium enterprises to access massive compute resources for short, semi-predictable time periods without having to build their own big data firms (Gantz & Reinsel, 2011).

The research methodology also includes comparing the results of a survey on human executive opinions against big data analytics results. For this purpose a questionnaire is prepared, the content of which is sourced from Deloitte's 2012 global out-sourcing and in-sourcing survey while designing the questionnaire.

In addition to it the predictive analytics, models of big data analytics when developed can be used to predict risk, fraud, opportunity and demand by answering questions like how risky is this deal? How fraudulent is this claim? What will maximize the customer profitability, where are the possibilities to reduce cost? How much demand is there likely to be in the future for this product or service? (Taylor, 2013).

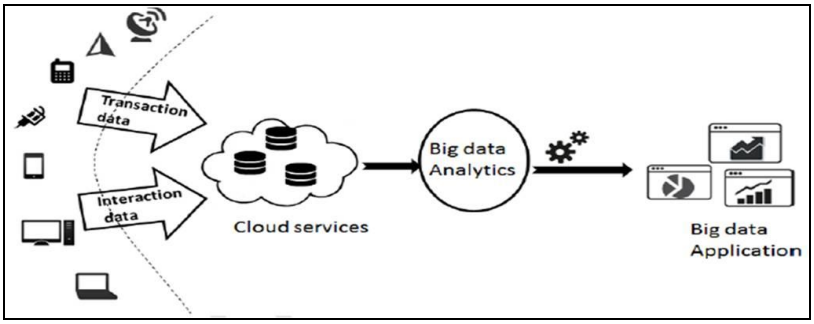

Figure 1a: Enterprise Decision Support Methodology through Big Data Analytics (Source: Neaga & Hao, 2013)

Figure 1b: Enterprise Management Methodology through Enterprise Data Warehousing & Data Mining (Source: Ferguson, 2012)

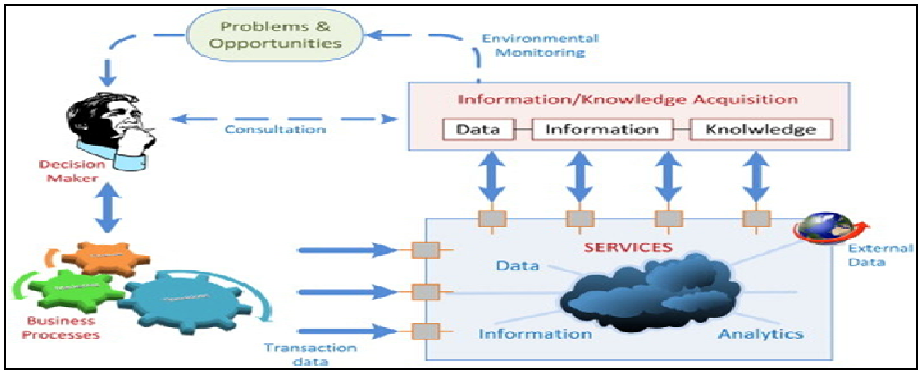

Figure 1c: Enterprise Decision Support Methodology (Source: Delen & Demirkan, 2013)

5. Big Data Analytics & Optimization

Big Data analytics is the application of advanced analytic techniques to very big (large) data sets. Big data is not just about size but also about diversity of data types and streaming data. Hence there are three 'V's of big data that constitutes the comprehensive definition. They are: volume, variety and velocity (Russom, 2011), in fact it becomes four V's when included the veracity characteristic too (IBM-SBM Forum, 2014).

Big Data Analytics describes the efficient use of a simple model applied to large volumes of data that would be too large for the traditional analytical environment (Revolution Analytics, 2011). Big data analytics can benefit customer relations, business intelligence and many analytic applications. Big Data Analytics potentially generates the outcome value of providing business opportunities but not as a solution to a problem. Insights on new customer segments or about cost reductions and providing the opportunities (Russom, 2011).

Research suggests that a simple algorithm with a large volume of data is more accurate than a sophisticated algorithm with little data. Algorithms do not bring competitive advantage by itself rather their ability to apply it to huge amounts of data – without

compromising performance and thus generates the competitive advantage. Increasingly analysis algorithms are provided directly by database management system (dbms) vendors. Analytics delivers optimized statistical algorithms for the three primary data management paradigms being employed to address growing size and increasing variety of organizations' data including file-based, MapReduce (e.g. Hadoop) or In-Database Analytics (Revolution Analytics, 2011). But as there is an exponential growth of data and its availability in the enterprises is observed, the importance of algorithms is highlighted as well as performing ETL for making the data base data cleansed in this article.

6. Theoretical Framework

In the last fifteen years, academic research on management information systems (IS) outsourcing has evolved rapidly. Indeed the field of outsourcing research has grown so fast that there has been scant opportunity for the research community to take a collective breath, and complete local assessment of research activities to date (Dibberm, Goles, Hirschhein & Jayatilaka, 2004). It is believed that the phenomenon of outsourcing as a case study of outsourcing experiences would benefit the organisations like King Saud University, Riyadh, Saudi Arabia regarding making decisions on outsourcing strategies.

For example, the study conducted by Di Giacomo & Brunzel (2010) advocates that combined IT outsourcing theories and additional concepts are able to provide a better output to the IT outsourcing decision process.

Further the business model suggested by AlSudairi & Vasista (2012) has actually suggested more number of theories to be referred and combined than what has been referred by Di Giacomo & Brunzel (2010) in this regard. The idea is that one can enable strategic decision based on the forensic & other data mining reports, real-time data mining analytics and predictive analytics (SNIA, 2012).

The outsourcing phenomenon is classified as tangible (IT outsourcing) and intangible (Knowledge outsourcing e.g. consultants). Scope of this research is limited to make decision on knowledge outsourcing. The survey is planned to conduct in King Saud University, Riyadh, Saudi Arabia. The respondents are targeted from the senior staff of the university. The responses will then be analyzed on using statistical data analysis on the collected data to examine how they are contributing towards knowledge management strategy of the university towards knowledge out sourcing.

A customized questionnaire is designed, which is sourced from Deloitte's 2012 global out-sourcing and in-sourcing survey for the purpose of getting insights on outsourcing scenario today and tomorrow. The demographic information is targeted to obtain information on respondent's continental location, nationality, department affiliation, specialization, annual revenue.

The questionnaire is prepared and available in APPENDIX-I at the end.

Effective outsourcing requires a clear understanding of the business drivers and goals, a commitment to seeing the effort through, careful planning and oversight, meticulous attention to detail and a lot of hard work. It is believed that this customized questionnaire will benefit King Saud University in generating an insight on knowledge outsourcing decisions.

The following section briefly describes environments and software tools that can provide the capabilities to optimize statistical algorithms under file based environment, with Map Reduce (e.g. Hadoop) and also with Database Analytics. It also describes the importance of data cleansing and protecting & retaining the semantics of big data.

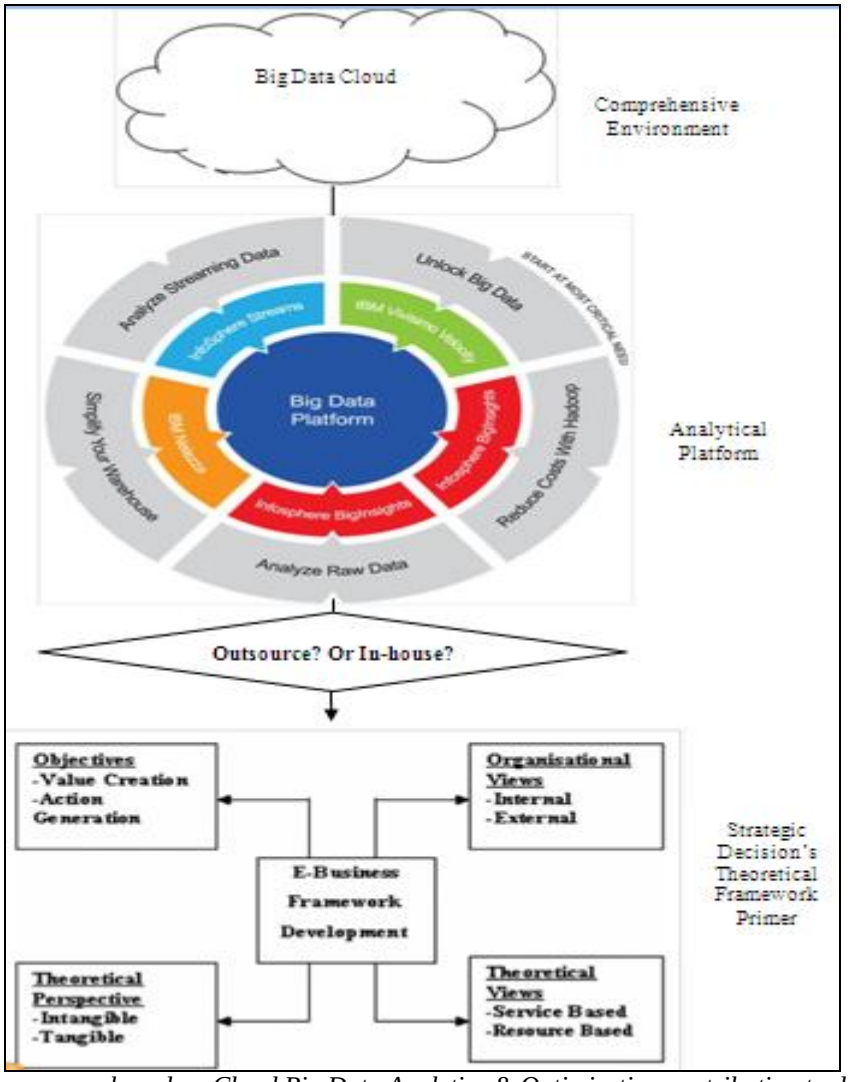

Figure 2: Decision process based on Cloud Big Data Analytics & Optimization contributing to decision making in Macro and Micro environments (e.g. outsourcing decision) (Note: Partially Big Data Platform Diagram is Sourced at Pavlik, 2013; & the strategic decision's theoretical framework is sourced at AlSudairi & Vasista, 2012)

A customized questionnaire is designed, which is sourced from Deloitte's 2012 global out-sourcing and in-sourcing survey for the purpose of getting insights on outsourcing scenario today and tomorrow. The demographic information is targeted to obtain information on respondent's continental location, nationality, department affiliation, specialization, annual revenue.
The questionnaire is prepared and available in APPENDIX-I at the end.
Effective outsourcing requires a clear understanding of the business drivers and goals, a commitment to seeing the effort through, careful planning and oversight, meticulous attention to detail and a lot of hard work. It is believed that this customized questionnaire will benefit King Saud University in generating an insight on knowledge outsourcing decisions.
The following section briefly describes environments and software tools that can provide the capabilities to optimize statistical algorithms under file based environment, with Map Reduce (e.g. Hadoop) and also with Database Analytics. It also describes the importance of data cleansing and protecting & retaining the semantics of big data.

6.1. Optimized statistical algorithms for Hadoop & File based Analytics

'R' is an interactive open-source implementation facility of a language that facilitates doing statistical analyses. It is a powerful statistical programming language that can carry out important analyses such as generalized linear and generalized additive models and mixed models of linear and non-linear models belonging to the disciplines of Statistics, Operations Research and Data mining etc. (Knell, 2014). 'R' can be applied to problems with research analysis. In general, activities involved in research problem analysis are: hypothesis declaration, exploring and understand the statistical data, use different statistical techniques, drill down on various dimensions and perform statistical analysis (Revolution Analytics, 2011). But one fundamental challenge here is that R is designed to work on memory-resident data only.

For example, KNIME and WEKA are the software tools to perform machine learning and data mining tasks. Its algorithms can be applied directly to a data set from its own interface. They usually contain tools for data pre-processing, classification, regression, clustering, association rules and visualization (Alcala-Fedez et al., 2009).
Programs in Hadoop (that uses map reduce technique) work independently and in parallel on individual data slices. A very large data set can be reduced into a smaller subset where analytics can be applied. The business value extraction logic is written as MapReduce jobs in Java. However using higher level languages like Hive (Apache Hive is a data warehouse software that facilitates querying and managing large data sets residing in distributed storage) and Pig (Pig is a high-level data-flow language for expressing Map/Reduce programs for analyzing large HDFS-Hadoop Distributed File System distributed data sets) can make writing these programs much easier. 'R' can be used to do the analysis on the output of MapReduced programs when written back to HDFS/HBASE and place on a data warehouse (Revolution Analytics, 2011). HBASE is an open-source, distributed, versioned, column-oriented store modeled after Google 'Bigtable' (Ferguson, 2012).
Zhang & Duchi (2013) have analyzed two communication-efficient algorithms for distributed optimization statistical setting involving large scale data sets. The first algorithm is a standard average method that distributes the N data samples evenly to m machines, performs separate minimization on each subset and then averages the estimates. The second algorithm is a novel method, based on an appropriate form of bootstrap sub-sampling. They have made an experimental evaluation on their proposed methods and found that the performance of both can efficiently solve an advertisement prediction problem. The performance is evaluated against the simulated data and on a large-scale regression problem from internet search domain (e.g. from the Chinese SoSo search engine).

6.2. Optimized statistical algorithms for In-Database Analytics
MADlib is a free, open-source library of in-database analytic methods. It provides an evolving suite of SQL-based algorithms for machine learning, data mining and statistics that run at scale within a database engine, with no need for data import/export to other tools. The goal of MADlib is to eventually serve as a role for scalable database systems that is similar to the CRAN library for R: a community repository of statistical methods that is written with scale and parallelism in mind (Hellerstein, 2012).
Conceptually the logic acts against considering the analytical attributes of the statistical databases as canonical coefficients that serves as parameters. These parameters can be used to solve the usual statistical queries with no access to actual data the values of which sits in the traditional database It is a kind of representation of a distributed data dictionary that contains the canonical coefficients of the domains of the relations in the Distributed Data Base (DDB) environment, which can permit an accurate forecast of the selectivity factor for the selection and join operations as well as by optimizing the AVERAGE and PERCENT query logics of algorithm (Lefons, Silverstri & Tangorra, 1983).

6.3. Optimized statistical algorithms for Hybrid Analytics
There are some challenges that are needed to be addressed relevant to the current context of the topic considered. They are: What can be done with large data sets that cannot be loaded and manipulated in main memory? Can abstract data-access primitives embedded in database systems provide mining algorithms with the information to drive a search for patterns? How to develop scoring models that focus on notions of accuracy and utility measures to derive the benefit of a specific pattern (such as money saved) (Fayyad & Uthurusamy, 2002).
For example, Oracle Big Data Connector is a software suite that integrates processing in Hadoop with operations in a data warehouse. It is designed to leverage the latest features of Apache Hadoop. Big Data Connectors connect Hadoop clusters with data infrastructure to harness massive volumes of both structured and unstructured data for critical business insights (Oracle Data Sheet, 2014).
Further the research study conducted by Neaga & Hao (2013) provides shell level knowledge on algorithms, methods and system relevant to big data analytics.

6.4. Cleansing data for mining and warehousing
Given the rapid growth of data, it is important to extract, mine and discover useful information from databases and data warehouses. Cleaning data from impurities is an integral part of data processing and maintenance (Muller & Freytag, 2003). The process of data cleansing is crucial because of the quote: GIGO-Garbage-In Garbage-Out principle. Data cleansing process deals with incorrect or missing data values, inconsistent value naming conventions, incomplete information and de-duplication of records (Li, Hongjun, Ling and Ko, 1999). Data cleaning and data transformations address major data quality problems. Data transformations are needed to support any changes in the structure, representation or content of data. These transformations becomes necessary for situations such as to deal with schema evolution (or to deal with master data management), migrating a legacy system to a new information system or when multiple data sources are to be integrated. In order to effectively deal with Data cleaning process, it passes through several phases such as (i) data analysis (ii) definition of transformation workflow and mapping rules, (iii) verification (conflict resolutions), (iv) transformation and (v) backflow of cleaned data. idCentric (FirstLogic), PureIntegrate (oracle), QucikAddress (QASSystems), Reunion (PitneyBowes) and Trillium (Trillium Software) are some of the special domain data cleaning tools. Tools such as DataClanser (EDD) Merge/PurgeLibrary Sagent/QMSoftware, IntelliClean are useful for de-duplication. Copy Manager(Information Builders), Data Stage(Informix/Ardent), Power Mart (Informatica), Decision Base (CA/Platinum), Data Transformation Service (Microsoft), Warehouse Administrator (SAS) are some of the Extract-Transform-Load (ETL) tools. Data cleaning is not only required for data warehousing but also for query processing on heterogeneous data sources (Rahm & Do, 2000).

6.5. Retaining and Protecting Semantics in Big Data

One of the major problems in big data analysis is to protect and retain the contextual relevance or semantics.

The increasing popularity of social network platforms has given opportunities to capture and understand the social structures among people and types of entities on the web. The research by Sun, Datta, Lim & Chang (2011) has demonstrated the innovative features of SSNetViz with social networks from three information sources covering a similar set of entities and relationships in terrorism domain. SSNetViz is developed for integrating, visualizing and querying heterogeneous semantic social networks obtained from multiple information sources. Mika (2004) presented the FLINK system for the extraction, aggregation and visualization of online social networks. FLINK system is targeted to employ semantic technology for reasoning with personal information extracted from a number of electronic information sources including web pages, e-mail publication archives and profiles. The acquired knowledge is use for the purpose of social network analysis and generating web-based presentation of the community. It has the capability to accurately trace the digital participation through ubiquitous and mobile means asking the governance the need and focus on standard representations, privacy and security of leading happy and non-problematic social life within digital society.

Ontology Visualization tools such as OntoVis for example makes the use of ontology information and several statistical measures to reveal the hidden knowledge in heterogeneous social networks with its powerful semantic and structural abstractions and filtering capabilities (Shen, Ma & Eliassi-Rad, 2006).

7. Current challenge in Cloud Big Data Analytics and addressing it

So it means there is a need for algorithms that helps end users in gaining insight from dynamic enumerated data by focusing on the extraction of patterns and turns them into meaningful reports and summaries by trading off complexity and understandability.

So the question is can Cloud Based Big Data Analytics provide effective decision making capability?

To some extent the answer is both 'YES' and a 'NO'.

It is 'YES' because for example, Oracle 'R' advanced analytics for Hadoop runs R code in a Hadoop cluster for scalable analytics. Oracle R advanced analytics for Hadoop accelerates advanced analytics on Big Data by hiding the complexities of Hadoop-based computing from 'R' end users. The connector integrates with Oracle Advanced Analytics for Oracle Database, to execute R and in-database Data Mining computations directly in the database. Oracle 'R' advanced analytics for Hadoop delivers faster insights by including a rich collection of high performance, scalable, parallel implementations of common statistical and predictive techniques, leveraging the Hadoop cluster without requiring data duplication or data movement. Thus connector enables analysts to combine all the data from several environments – client desktop, HDFS, HIVE, Oracle database and in-memory 'R' data structure- all in the context of a single analytical task execution (Oracle Data Sheet, 2014) where Hadoop HDFS is a distributed file system that partitions large files across multiple machines for high-throughput access to data; and HIVE is a data warehouse system for Hadoop that facilitates data summarization, ad-hoc queries and the analysis of large datasets stored in Hadoop compatible file systems. HIVE provides a mechanism to project structure onto this data and query it using a SQL-like language called HiveQL. HiveQL programs are converted into Map/Reduce programs.

It is 'NO' because there are no automated generic capabilities that can handle and can work on context based dynamic enumerated data out of which patterns and insights can be retrieved. The only way for complete 'YES' is the synergetic combination of automated tools, software tools and the skills from human staffing.

So, practically speaking one cannot expect OLAP to go away. Online Analytical Processing (OLAP) is by far the most common analytic approach today especially to encounter the problem of big data that contain the enumerated dimensions and to generate analysis on such type of data. Hence it demands adequate staffing or skills, a lack of business support and problems with using database software and analysis software tools. These skills also include the management of governance, privacy and security of enterprise data (Russom, 2011).

7.1. Addressing the Cloud Big Data Challenge

Addressing such kind of requirement is obvious because many enterprises are choosing hybrid strategies in which they're moving selected big-data workloads into the cloud while keeping sensitive data and mission-critical workloads on-premises. So the answer is a kind of 'YES' with Amazon Web Services. Amazon Web Services (AWS) after having partnered with the leading Cloudera, apparently reasoning that many enterprises are choosing hybrid strategies in which they're moving selected big-data workloads into the cloud. AWS even extends this connectivity to the enterprise data centers level (Information Week, 2014 Jan 30) or with the use of similar kind of environment such as with IBM and other leading vendors. Rather than providing a comprehensive tool, IBM relies on building comprehensive strategy of solving enterprise management problems (Russom, 2011).

7.2. Ten important cloud big data analytic tools

The ten important cloud big data analytics platforms that can help in deriving patterns and insights are: (i) Amazon Web Services (ii) Cloudera (iii) HP HAVEn (iv) IBM (v) MapR (vi) Microsoft (vii) Oracle (viii) Pivotal (ix) SAP and (x) Knowledge STUDIO.

Some of the evaluating factors of Cloud Big Data Analytics Performance include: the capabilities of Analytical DBMS, In-memory DBMS, Hadoop distributions, Stream-processing technology and Hardware & Software systems (Information Week, 2014).

8. Conclusion

Cloud Big Data Analytics provide the new approach needed to produce efficient and affordable applications for discovering knowledge and patterns from very large/big data sets directed to support business intelligence and decision support system applications. Analytics-as-a-Service can be evolved under cloud business paradigm. Cloud based business analytics is also cost effective, easy to set up and test.

Supply chain performance is directly related to cost efficient velocity. It is a realized scenario that vertical Integration up and down the supply chain between trading partners has generated significant benefits even with on-premise technology. However if the scenario has to upgrade to fit the peer-to-peer collaboration, such process cannot effectively managed by on premise tools from any single participant; it can be managed only from a neutral, cloud platform-based perspective (One Network Enterprise, 2014).

The concept of Experimentation-as-a-Service (EaaS) constitutes a move from the current model of federation-by-design, to ad hoc, on demand federation and reconfiguration of facilities and platforms, with potentially different purposes and focuses which were not initially supposed to be federated, in order to meet the researchers' or experimentation needs (EC 2020 Initiative, Undated).

9. Future Research

It is expected that Universities such as King Saud University, Riyadh, who are blessed with the highest budget awards from His Excellency and the custodian of two holy mosques and the president of Kingdom of Saudi Arabia will put an effort in experimenting to derive insights and decision patterns related to out-sourcing using cloud big data analytics approach so as to help the nation drive towards achieving knowledge based economy. The results obtained out of this experimental effort can be compared with the results of survey mentioned in APPENDIX-I for its compatibility and usefulness of making decisions and building strategies.

10. Appendix-I

Administratively, chairs and above, academically associate professors and above are requested to participate in the survey.

(A) Demographics Section

1. I belong to: (a) Asia (b) Europe (c) Africa (d) Australia (e) North America (f) South America continent
2. My nationality is: _____
3. I work for the _____ department.
4. My specialization in educational qualification is: ----------
5. My specialization in professional qualification is: ----------
6. My annual revenue is: ----------
7. My Total Number of Years of Experience: ----------

(B) Knowledge Outsourcing Questionnaire Section

1. Is outsourcing a standard operating practice at your college?
2. Which department or specialization is currently functioning under outsourcing option?
3. Which department or specialization is planning to go for outsourcing?
4. How supportive is outsourcing to the college functioning? (a) Supportive (b) unsupportive
5. We are hiring contracted people for the ---------- department of our college
6. Majority of outsourcing is expected to occur in ---------- department of our college
7. How important were each of these objectives in your most recent outsourcing effort?

Specify (a) Very important (b) important (c) Not important for the following:

(i) Reduce operating costs (ii) improve customer service (iii) gain competitive advantage (iv) leveraging new and innovative knowledge (v) Access to flexible teaching practices (vi) Access to flexible learning practices (vii) desire to have consolidation (viii) Improve controls

8. Savings of Knowledge outsourcing (a) exceeded expectations (b) met expectation (c) fell short of expectations (d) feeling not a viable solution
9. How satisfied are you with knowledge out sourcing from your most recent initiative

(a) Extremely satisfied (b) Satisfied (iii) Neutral (iv) Dissatisfied

10. If your option is either Neutral or Dissatisfied please answer the following:

(i) Consultant is underestimated the scope/effort
(ii) Lack of service level attainment
(iii) Low consultant professional performance
(iv) Lack of timely project/service request execution
(v) Attrition of key resources
(vi) Lack of innovative efforts
(viii) Lack of domain functional knowledge
(ix) Lack of overall service quality
(x) Lack of subject matter expertise
(xi) Pricing versus performance

(xii) Unsuccessful transition
(xiii) Communication
(xiv) Cultural fit

11. If you are not satisfied with current out-sourcing phenomenon what do you want to do?
(a) Move to other (b) in-source
12. If you have opted for post in-source option, how satisfied are you with post-in-sourcing
(a) Extremely satisfied (b) Satisfied (c) Neutral (d) Not satisfied

(C) IT Outsourcing Questionnaire Section
1. Adopting of Cloud computing act as a driver to our department
(a) Yes (b) No (c) Don't know
2. Is your college (of a university) currently using cloud computing services?
3. Which component do you feel that it should become part of cloud?
(i) e-mail messaging (ii) web site hosting and maintenance (iii) IT infrastructure as a service (iv) software platforms as a service (v) application packages as a service (vi) front of systems (vii) back office systems (viii) voice/telephony systems (ix) middle office systems
4. How important is cloud computing service for your college? (a) Very important (b) important (c) not important
It reduces the capital spending; it increases the agility; it reallocates resources; it allow access to new technologies or services
5. Do you plan to use cloud computing within 3 years?
(a) Yes (b) No (c) Don't know
6. IF No, why do you believe you will not use cloud computing?
(i) Cloud computing is too new a concept for my college business; (ii) Data Privacy concerns; (iii) security concerns (iv) my college business does require the knowledge to implement cloud computing (v) my college business do not have the expertise and knowledge to deal or implement cloud computing.
7. How satisfied are you with IT out sourcing from your most recent initiative
(a) Extremely satisfied (b) Satisfied (iii) Neutral (iv) Dissatisfied
8. If your option is either Neutral or Dissatisfied please answer the following:
(i) Vendor/consultant is underestimated the scope/effort
(ii) Lack of service level attainment
(iii) Low vendor/consultant professional performance
(iv) Lack of timely project/service request execution
(v) Attrition of key resources
(vi) Lack of innovative efforts
(viii) Lack of domain functional knowledge
(ix) Lack of overall service quality
(x) Lack of subject matter expertise
(xi) Pricing versus performance
(xii) Unsuccessful transition
(xiii) Communication
(xiv) Cultural fit
9. If you are not satisfied with current IT out-sourcing phenomenon what do you want to do?
(a) Move to other (b) IT in-source
10. If you have opted for post in-source option, how satisfied are you with post IT in-sourcing
(a) Extremely satisfied (b) Satisfied (c) Neutral (d) Not satisfied

11. References

1. Alcala-Fedez, J. et al. (2009). KEEL: a software tool to assess evolutionary algorithms for data mining problems. Soft Computing, 13: 307-318, Springer.
2. AlSudairi, M. A. T. & Vasista, T. G. K. (2012). Model for value creation and action generation of an electronic enterprise in a knowledge based economy. International Conference on Information Society (i-society). June 25-28, London, UK.
3. AlSudairi, M., Dwivedi, Y. K. (2010). A Multi-disciplinary profile of IS/IT outsourcing research. Journal of Enterprise Information Management, 23 (2), 215-258.
4. Anonymous (Undated). Transaction Cost Economics of Cloud Computing Adoption. URL: http://vslir.iimahd.ernet.in:8080/xmlui/bitstream/handle/123456789/11613/TIK-PP-246-TRANSACTION_COST_ECONOMICS_OF_CLOUD_COMPUTING_ADOPTION-165-Gunupudi_b.pdf?sequence=3 (Accessed on 08-09-2014).
5. Deloitte (2012). Outsourcing, today and tomorrow: Insights from Deloitte's 2012 global out-sourcing and in-sourcing survey. URL:

http://www.deloitte.com/assets/Dcom-Malaysia/Local%20Assets/Documents/Consulting/Strategy%20and%20operations/insights_global_outsourcing_insourcing_survey_2012.pdf (Accessed on 09-09-2014).

6. Delen, D. & Demirkan, H. (2013). Data, Information and analytics as services. Decision Support Systems, 55 (1), 359-363.
7. Demirkan, H. & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. Decision Support Systems, 55(1), 412-421.
8. Di Giacomo, D. & Brunzel, T. (2010). Cloud Computing Evaluation – How it differs to Traditional IT outsourcing. Master Thesis within IT & Management. Jonkoping International Business School, Jonkoping University.
9. Dyer, J. H. & Singh, H. (1998). The Relational View: Cooperative Strategy and Sources of Inter-organisational Competitive Advantage. Academy of Management Review, 23(4), 660-679.
10. Dibbern, J., Goles, T., Hirschheim, R., & Jayatilaka, B. (2004). Information systems outsourcing: a survey and analysis of the literature. ACM SIGMIS Database, 35(4), 6-102.
11. EC 2020 Initiative (undated). Digital Agendra for Europe, A Europe 2020 Initiative: From Experimentation to Future Internet. URL: http://ec.europa.eu/digital-agenda/en/experimentation-future-internet (Accessed on 09-09-2014).
12. Fayyad, U. & Uthurusamy, R. (2002). Evolving Data Mining into solutions for insights. Communications of the ACM, 45 (8), 28-31.
13. Ferguson, M. (2012). Architecting a Big Data Platform for Analytics. A White Paper by Intelligent Business Strategies, Cheshire, SK9 5BG, England, UK.
14. Foss, N. J. (2003). The strategic management and transaction cost nexus: past debates, central questions, and future research possibilities. Strategic Organization, 1(2), 139-169.
15. Gantz, J. & Reinsel, D. (2011). Extracting Value from Chaos. IDC IVIEW. This paper uses bits and pieces of the excerpts of this reference sponsored by EMC.
16. Hellerstein, J. M. et al. (2012). The MADlib analytics library: or MAD skills, the SQL. In Proceedings of the VLDB Endowment, 5(12), 1700-1711, ACM Digilib.
17. IBM-SBM Forum (2014). IT infrastructure matters because business outcomes matter. 14th September, Riyadh Mariott Hotel, Riyadh, Saudi Arabia
18. Information Week (2014, Jan, 30). 16 Top Big Data Analytics Platforms. URL: http://www.informationweek.com/big-data/big-data-analytics/16-top-big-data-analytics-platforms/d/d-id/1113609?image_number=4 (Accessed on 08-09-2014).
19. Knell, R. (2014). A beginner's guide to data visulaisation and statistical analysis and programming in R. ISBN 978-0-9575971-1-2
20. Lefons, E., Silvestri, A., & Tangorra, F. (1983, October). An Analytic Approach to Statistical Databases. In VLDB, 260-274.
21. Li, M., Hongjun, L., Tok, L., Ling, W. and Ko, Y. T. (1999). Cleansing data for mining and warehousing. In the proceedings of 10th Database and Expert Systems Applications (DEXA), August30-September 3, Florence, Italy.
22. Melville, N. Kraemer, K. & Gurbaxani, V. (2004). Review: Information technology and organisational performance: an integrative model of it business value. MIS Quarterly, 28 (2), 283-322.
23. Mika, P. (2005). Web Semantics: Science, Services and Agents on the World Wide Web. Selected Papers from the International Semantic Web Conference-ISWC, 2004, 3(2-3), 211-223.
24. Muller, H. & Freytag, J-C. (2003). Problems, Methods and Challenges in Comprehensive Data Cleansing.
25. Ndlela, L. T. & du Toit, A. S. A (2001). Establishing a knowledge management programme for competitive advantage in an enterprise. International Journal of Information Management, 21 (2), 151-165.
26. Neaga, I. & Hao, Y. (2013). The Business Impact of Cloud based Big Data Analytics. URL: http://ec.europa.eu/digital-agenda/events/cf/ict2013/document.cfm?doc_id=26161 (Accessed on 09-09-2014)
27. Nickerson, J. A. & Silverman, B. S. (1997). Integrating Competitive Strategy and Transaction Cost Economics: An Operationalization of Fit in the Interstate Trucking Industry. Washington University Working Paper OLIN-97-06. Available at SSRN: http://ssrn.com/abstract=79949 (Accessed on 08-09-2014).
28. One Network Enterprise (2014). Cloud-Based Supply Chain Technology vs. Traditional On-Premise: Tech Brief Cloud-Based Supply Chain. URL: Cloud Versus Traditional Supply Chain Technology (PDF) (Accessed on September 14, 2014).
29. Oracle Data Sheet (2014). Oracle Big Data Connectors. URL: http://www.oracle.com/technetwork/database/database-technologies/bdc/big-data-connectors/overview/bigdata-connectors-datasheet-1883359.pdf (Accessed on 08-09-2014)
30. Pavlik, M. (2013). IBM Big Data platform Overview. URL: http://www.ibm.com/systems/data/flash/hu/Bigdata/4_exabyte/1_Mi_is_az_a_big_data_MartinPavlik.ppt (Accessed on 08-09-2014).
31. Powell, T. C. & Dent-Micallef, A. (1997). Information Technology as competitive advantage: The role of human, business and technology resources. Strategic Management Journal, 18(5), 375-405.
32. Rahm, E., & Do, H. H. (2000). Data Cleaning: problems and Current Approaches. IEEE Bulletin on Data Engineering, 23(4), 3-13.
33. Revolution Analytics (2011). Advanced 'Big Data' Analytics with R and Hadoop. A white paper from Revolution Analytics. URL: http://www.revolutionanalytics.com/sites/default/files/r-and-hadoop-big-data-analytics.pdf (Accessed on 08-09-2014).

34. Russom, P. (2011). Big data analytics. TDWI Best Practices Report, Fourth Quarter. URL: http://public.dhe.ibm.com/common/ssi/ecm/en/iml14293usen/IML14293USEN.PDF (Accessed on 08-09-2014).
35. SNIA (2012). Introduction Big Data Analytics _ Hadoop. URL: http://www.snia.org/sites/default/files2/ABDS2012/Tutorials/RobPeglar_Introduction_Analytics%20_Big%20Data_Hadoop.pdf (Accessed on 08-09-2014).
36. Shen, Z., Ma, K-L & Eliassi-Rad, T. (2006). Visual Analysis of Large Heterogeneous Social Networks by semantic and Structural Abstraction, IEEE Transactions on Visualization and Computer Graphics, 12(6), 1427-1439.
37. Sultan, N. A. (2011). Reaching for the ''cloud'': How SMEs can manage. International Journal of Information Management, 31(3), 272-278.
38. Sun, A., Datta, A., Lim, E-P & Chang, K. (2011). Visualizing and Querying Semantic Social Networks. In the Proceedings of SIGIR'11, July 24-28, Beijing, China, ACM
39. Taylor, J. (2013). Predictive Analytics in the Cloud: Opportunities, Trends and the Impact of Big Data 2013. Decision Management Solutions, URL: http://decisionmanagementsolutions.com/attachments/article/294/PACloud2013ReportFinal.pdf (Accessed on 09-09-2014).
40. Valencia & Clemmons (2011-2012). Idea in brief and in practice provided on the original article 'Putting the Enterprise into the Enterprise System' by Davenport T. H. (1998)'s Putting the Enterprise into the Enterprise System. Harvard Business Review. URL: http://people.stfx.ca/x2009/x2009kap/448%20Papper's%20ref/83837_1_Putting-the-Enterprise-Into-the-Enterprise-System-1-.pdf (Accessed on 08-09-2014).
41. Zhang, Y. & Wainwright, M. J. (2013). Communication-Efficient Algorithms for Statistical Optimization. Journal of Machine Learning, 14, 3321-3303