Dark data: Business Analytical tools and Facilities for illuminating dark data

[1]Njeru Mwiti Kevin,
njerukevin@gmail.com

[2] Felister Munyi Wanyaga,
Thika technical training Institute, Thika, Kenya.
ferliem2011@gmail.com

[3] David Kibaara
kibaaragichuiri@gmail.com

[4] Wilkister Atieno Dinda
School of Computing and Informatics, Mount Kenya University
wdinda2013@gmail.com

James Kariuki Ngatia; Safaricom Limited
jamesngatia@gmail.com

***Abstract*- The most important asset for any organization today is data. Organizations collect and store vast amounts of data every day relating to their various business activities. Understanding this data leads to better insights, less costs and risks and provides avenues in which the organization can improve its performance, offer better services to its customers and earn more revenue giving it a competitive advantage in the market. Advanced tools have been developed to gain this much needed insight into data that was previously considered irrelevant or inaccessible based on its unstructured form. These tools help an organization drill into its data and data from other external sources such as competitors, government reports, proprietary and other multi dimensional databases available from the internet to gain knowledge that can be applied to improve the organization's competitive position. The aim of this research is to provide insights to organizations on how business analytic tools and software can be applied in lighting up previously unknown or ignored data. This is done through an in-depth analysis of secondary data and practitioner reports to provide an understanding of the various concepts and tools essential in identifying meaningful patterns and trends into an organization's data.**

***Index Terms*— Dark data, knowledge, intelligence, analytics, competitive advantage, Big Data**

I. INTRODUCTION

According to [1] [2], 16 Petabytes of data are captured every day globally (Big data). As cited by [2], The McKinsey Global Institute estimates that data volume is growing 40% per year, and will grow 44x between 2009 and 2020. For example, 2.72 Zettabytes (ZB) of data was captured in 2012 alone, and this amount continues to double every year to reach 8 ZB by 2015 [1]. These arguments are supported by [2], who predicts that the amount of data being collected will grow tenfold from 4.4 trillion Gigabytes to 44 trillion Gigabytes by 2016.

It is estimated that 80% of this data is unstructured [3]. This means that much of the data stored within organizations has no benefits to the organization in its current state. This has been contributed by various factors including: IT storage optimization, system migration, compliance obligation and the global reduction in the cost of storage, meaning that organizations can now be able to store too much data and in diverse forms at a very low cost. Advancement in data capture devices that have enhanced the capture of unstructured data such as voice, audio, video and graphics has also created a large pool of data within and outside organizational boundaries. Perhaps this is explained more clearly by [2] in their definition of big data who describe big data as:

- Traditional enterprise data includeing customer information from a firm's Customer Relationship Management systems (CRM), transactional Enterprise Resource Planning (ERP) data, web store transactions, and general ledger data.
- Machine-generated /sensor data – includes Call Detail Records ("CDR"), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), and trading systems data.
- Social data – includes customer feedback streams, micro-blogging sites like Twitter, social media platforms like Facebook

The big data Concept

Companies capture trillions of bytes of information about their customers, suppliers, and operations, and millions of networked sensors are being embedded in the physical world in devices such as mobile phones and automobiles, sensing, creating, and communicating data [8]. Big Data is defined by [1]as huge data sets that are orders of magnitude larger (volume); more diverse, including structured, semi-structured and unstructured data (variety) and arriving faster (velocity) than any organization has had to deal with before. [9]defines big data as data that cannot be processed using traditional data processing tools and processes; while [8] defines "big data" as large pools of data that can be captured, communicated, aggregated, stored, and analyzed; [10] suggests that big data is a collection of data sets that are so large and complex that software systems are hardly able to process them.

Four key characteristics define big data [2], these are

Volume

Machine-generated data is produced in much larger quantities than non-traditional data. For instance, a single jet engine can generate 10TB of data in 30 minutes. With more than 25,000 airline flights per day, the daily volume of just this single data source runs into the Petabytes. Smart meters and heavy industrial equipment like oil refineries and drilling rigs generate similar data volumes, compounding the problem.

Velocity

Social media data streams – while not as massive as machine-generated data – produce a large influx of opinions and relationships valuable to customer relationship management. Even at 140 characters per tweet, the high velocity (or frequency) of Twitter data ensures large volumes (over 8 TB per day).

Variety

Traditional data formats tend to be relatively well defined by a data schema and change slowly. In contrast, non-traditional data formats exhibit a dizzying rate of change. As new services are added, new sensors deployed, or new marketing campaigns executed, new data types are needed to capture the resultant information.

Value

The economic value of different data varies significantly. Typically there is good information hidden amongst a larger body of non-traditional data; the challenge is identifying what is valuable and then transforming and extracting that data for analysis.

These vast amounts of data are generated by interconnected devices –from PCs and smart phones to sensors such as Radio frequency identification (RFID) readers and traffic cams. The growth of ubiquitous computing has also had profound effect on the gathering of data. Household devices such as fridges, microwaves and even digital television sets can gather data in their environment, analyze it and effectively provide relevant information to their users. This datais also heterogeneous and comes in many formats, including text, document, image, video, and more. Big data reflects not just how an organization identifies, analyzes and uses the data managed within its own walls, but also data that was previously considered inaccessible, including data from new sources of information that may lie outside the control of an organization, to make business decisions [11].

The Big Data phenomenon creates tremendous opportunities for society to foster data-driven innovation that can enable faster and better decision-making, building a competitive advantage for our knowledge society [1].When big data is distilled and analyzed in combination with traditional enterprise data, enterprises can develop a more thorough and insightful understanding of their business, which can lead to enhanced productivity, a stronger competitive position and greater innovation – all of which can have a significant impact on the bottom line [12]. To derive real business value from this data, an organization needs the right tools to capture and organize a wide variety of data types from different sources, and to be able to easily analyze it within the context of all your enterprise data. This is made difficult largely because of the unstructured form in which this data exists in and by the fact that not all this data is relevant to the firm. The first task therefore for any firm is to make sure that whatever data the firm stores is relevant to the enterprise. Various tools exist in the market that helps a company acquire and organize their data and also dig into this vast data to find meanings and relationships that are relevant to the firm. A good example is Oracle big data appliance and oracle big data connectors' tools that provide a complete and integrated solution to address the full spectrum of enterprise big data requirements [12].

To make the most of big data, enterprises must evolve their IT infrastructures to handle these new high-volume, high-velocity, high-variety sources of data and integrate them with the pre-existing enterprise data to be analyzed [2]. If this is not done, then enterprises may lose the overall benefit that would have been generated from this data.

II. Dark Data

Dark data is defined by Gartner[13] as "As the information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes (for example, analytics, business relationships and direct monetizing)." This definition is even made simpler by [14]who describe it as data that business and industry are paying to store, protect and manage yet it isn't being efficiently utilized to improve the value of their business.According to research by the Compliance, Governance and Oversight Counsel, 69% ofa company's stored data has absolutely no value to the organization. So, why keep it? They thus propose the use of content-based retention policies that will empower you to keep only datathat is important to the business hence gaining more benefits.

According to [14], keeping only content with business value can reduce retention costsby up to 70 percent. [14]Identifies five approaches that can be used to unveil dark data: Managing the growth of storage: this ensures that the organization keeps only data that has value to the business;Holistic capture of data: which ensures that all data produced within the organization is collected and stored

Deliver self-service access; because each group of users in your organization require different type(s) of data to support their duties, giving the users self-service to search for the data they need themselves from the organization's archives because they are the ones who know what they need most. This maximizes workforce productivity with transformational business insights into the stored data.Automation of the data lifecycle: establishing set governance policies for defensible content deletion can be a significant key in the reduction of "dark data." Defining a clear lifecycle for the data from the time it is created to when it is deleted. This helps not only to automate records retention but to also manage your data more strategically for improved compliance and litigation readiness. Assuring compliance and discovery: this helps to reduce costs and risks with enterprise-wide search and the efficient discovery of all Electronically Stored Information (ESI). It also ensures that the discovered information is simple and defensible and makes it easier for legal and compliance requirements.

[3] identifies four stages of lighting up data in an organization: *Identification; Classification of data; Controls and Continuous monitoring*. The Identification phase tries to find out what data the organization has and where it is stored (either within or outside the organization); classification organizes the data into groups to reflect the structural organizational needs or processes within the organization; the control stage manages the data that has been organized to ensure its security and integrity. This stage also introduces data analysis which ensures that the organization is able to gain insights into the data for business intelligence. Continuous monitoring ensures that proper mechanisms are put in place to ensure that the data is continually maintained to serve the needs of the organization. This stage also ensures return on investments (ROI) and identifies areas of improvements. This captures the overall concept of illuminating dark data using analytics (manage) because much of the data within organization is usually unstructured and can only be made usable if we are able to mine it and extract useful patterns and relationships which would the firm into making the best business decisions.

According to [2], companies will continue to waste 80% of customer data they have collected. Even more, the international data corporation (IDC) estimated that in 2013, only 22% of all the data in the world was useful (could be analyzed), and even in that case, only 5% was actually analyzed. A University of Texas study put these general estimates in a business context: it found out that for the medium Fortune 1000 company, a mere 10% increase in the usability of its data translates to an increase of $2.01 billion in annual revenues and a 10% increase in remote accessibility to data translates into an additional $65.67 million in net income per year [2].It should be noted that data that cannot be analyzed is useless to an organization.

On the relationship between dark data and big data, [15] opines that the problems brought about by "big data" start right away during data acquisition, when the data tsunami requires us to make decisions, currently in an ad hoc manner, about what data to keep and what to discard, and how to store what we keep reliably with the right data about data. This means that the dark data problems can only be solved during data acquisition. Where the organization needs to identify the requisite data to keep and what data adds no value to the organization. It is thus almost impossible to separate these two issues, reason being that it is the problem of big data that gives rise to dark data.

The biggest challenge for firms today is how to handle big data [8]; This is because these firms have access to a wealth of information which they don't know how to get value from as it is sitting in its raw form in either a semi-structured or unstructured state and as a result, they don't even know whether it is worth keeping [9]. They further argue that organizations are facing massive amounts of data; and organizations that don't know how to manage this data are overwhelmed by it.

Data-Driven Business decision making

The promise of data-driven decision-making is now being recognized broadly [15]. The goal of business decision making is to facilitate high revenue collection, lower the cost of production and lower the overall expenses incurred by the business [16]. [2]suggests that companies that excel in data driven decision making are 5% more productive and 6% more predictable than their competitors. Further, a study by IDC found that users of big data and analytics who collect their data from diverse sources and apply diverse analytic tools and metrics are five times more likely to exceed expectations in their projects than those who don't.

Therefore, a firm's decision making should be driven by an analysis of its internal and external environment. The benefits of using computers in this case is that they enable the firm to maximally collect data internally based on an organization's processes and transaction processing activities and at the same time link the firm with its external environment where Competitive intelligence (CI) data can be easily gathered through market analytics, trend analysis as well as secondary data sources from government institutions and other reliable sources outside the firm's boundaries.

Understanding Business Analytics

Business analytics is defined by [17] as project management tools and applications and data warehouse platform software used to access, transform, store, analyze, model, deliver and track information to enable fact-based decision-making and extend accountability by providing all decision-makers with the right information, at the right time, using the right technology. Business Analytics is also defined by (Techopedia) as computer software solutions that help users tap into enterprise data to make better, more informed business decisions. It therefore means that BI focuses on providing insights to a firm based on data gathered from within and outside a firm's boundaries; this data can be extracted from publicly accessible databases and

Through use of business analytics solutions, organizations can instantly identify the factors that impact their performance, create more accurate forward-looking strategies, enhance efficiency, increase profitability, and improve customer satisfaction and loyalty. [9] opines that vast amounts of data continue to overwhelm some organizations, yet opportunities exist with the right technology platform to analyze almost all kinds of data to gain a deeper understanding of the organization's customers, employees, products, competitors and the market place.

The use of BI for example is a strategy that can bring together the forces that drive business operations, i.e., people, processes, and technology in a collaborative environment and enable successful adoption of IS to deliver true business value [18]. Predictive analytics, social media analytics, data visualization, data mining, enterprise search, and location analytics and mapping solutions are among the many tools available in the business analytics category. Business analytics is all about the decisions that go into running a business [16].

Through the use of business analytics, "users can retrieve, combine, and explore data from a wide array of internal and external sources to uncover and identify patterns, trends, relationships, and anomalies. Business analytics can also enable users to conduct accurate forecasts, anticipating future events or results based on historical information."

[19] further argues that business analytics can hold the key to optimization of performance, informed decisions, actionable insights and trusted information by bringing together all relevant information in an organization thus companies can answer critical questions such as "What is happening?, Why is it happening?, What is likely to happen in the future?, and how to plan for that future?"[19]. They identify four components of business analytics namely: business intelligence; analytic application, advanced analytics and financial performance and strategy management.

Business analytics therefore helps an organization make better decisions through business insights, planning and performance by unlocking data captured in operational and financial systems and then transforming it into more useful and relevant information. Business analytics technologies can be placed under the umbrella of Enterprise Decision Management (EDM) systems [20]. These systems exploit methods and technologies to improve the efficacy and efficiency of decision making through the organization. The operational comparison of EDM and the traditional computer automation are described by Butler analytics (2014) as shown in Figure1.

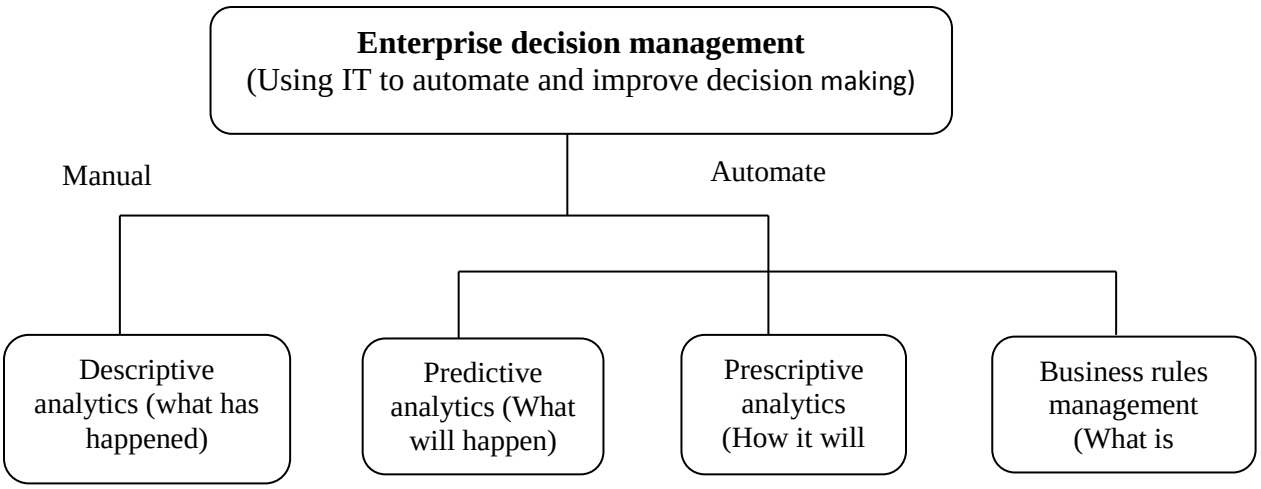

Figure 1: Comparison between EDM and Automation

This demonstrates the benefits of using EDM systems in organizations as opposed to the traditional computer driven automation of organizational functions. Electronic decision management systems act as uncertainty reduction machines, employing the use of statistics, machine learning, data mining, optimization and business rules engines to fine tune decisions and increase massively the speed at which decisions are made [20].

Business Intelligence

Business Intelligence (BI) combines data, analytical tools, methodologies, and new information extracted from data, with business knowledge, and targets them into decision making process [21]. BI is defined by [10] as a set of processes and technologies that convert data into meaningful and useful information for business purposes. BI is focused on querying and reporting. The difference between Business intelligence and business analytics is that, BI seeks to answer questions such as what is happening now and where, and at the same time seeks to explain what business actions are needed based on prior experience. Business Analytics, on the other hand, seeks to answer questions like why something is happening, what new trends may exist, what will happen next, and what is the best course for the future.

The role of BI in organizations is stressed by [18] who opines that a BI strategy aligns with the enterprise goals, improves knowledge management, advances business by making the best use of information, enables BI penetration into the business processes and helps an enterprise with strategic, tactical, and operational decision making.

BI can have a direct positive impact on the business performance of an enterprise, dramatically improving the ability to accomplish the mission by making smarter decisions at every level of the business from corporate strategy to operational processes. BI systems combine different data resources into information about processes in the company and provide this information in appropriate way and in a timely manner tosupport a company's management decision making. The information is usually based on analysis of enterprise operational data and serves as a basis for tactical and strategic decisions.

Every organization desires to find new ways to improve performance, develop stronger customer relationships, grow revenue and increase workforce effectiveness. BI has been considered a critical factor in achieving such results because of its ability to support informed decision making at every level; enabling managers, executives and knowledge workers to take the most effective action for every situation [22]. With Business Intelligence, companies can support decisions by establishing a fact-based decision making framework through a strong computer system that provides confidence in any decisions made [23].

Business intelligence (BI) is the combination of tools, processes, and skills that help turn the vast amount of a firm's data into digestible information [2].BI can have a direct positive impact on business performance of an enterprise by dramatically improving the ability to accomplish the mission, by making smarter decisions at every level of the business from corporate strategy to operational processes.

BI strategy

For BI to be successful, a firm should first of all build an effective BI strategy, which is driven by business objectives, enables stakeholders with better decision making capabilities and helps enterprise achieve desired goals [18]. An effective BI strategy should ensure that enterprise objectives, business strategy, investments, and BI are aligned. Enterprises that are able to connect BI to overall enterprise objectives become intelligent enterprises.BI strategy that aligns with the enterprise goals, improves knowledge management, advances business by making the best use of information, enables BI penetration into the business processes, and helps enterprise with strategic, tactical, and operational decision making.BI strategy brings together the forces that drive business operations, i.e., people, processes, and technology in a collaborative environment and highlights the approach to enable successful adoption of BI to deliver true business value [14].

The purpose of building a BI strategy is to help a business with long-term planning, help middle management with tactical reporting, and help operations with day-to-day decision making to run the business efficiently. BI is aims at providing people with the information they need to do their jobs more effectively.

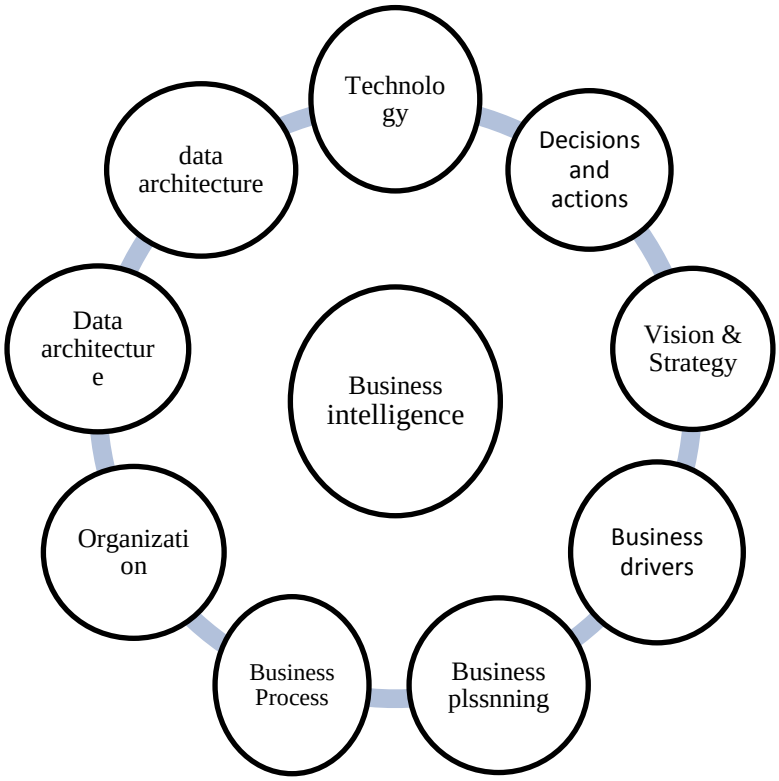

Figure 2: BI alignment with business strategy

Steps to ensuring successful BI strategy:

[14] proposes the following steps as key to ensuring the success of a BI strategy:

i. Create a business case and outline the expected benefits.

ii. Obtain buy in from stakeholders, especially the senior executives.

iii. Have an enterprise-wide perspective.

iv. Establish a criterion for success

v. Treat data and information as organizational asset.

vi. Adopt best practices and standards.

vii. Set up change management procedures.

viii. Business Intelligence strategy should align with the overall IT strategy and enterprise goals.

ix. Do a current state, future state, and gap analysis.

x. Think actionable and in small reliable steps.

xi. Establish a governance body.

xii. Use iterative implementation approach with parallel tracks

xiii. Work with designed frameworks and adopt proven methodologies

xiv. Assess BI readiness of the organization and identify related gaps and issues.

xv. Analyze and document the constraints and assumptions

xvi. Consider all the components of BI.

Role of business analytics in lighting-up dark data

In 2010, [24], predicted that in a few years, data would come at organizations in every form imaginable. Video, text, email, conversations, photos will dominate our databases in the future. Systems that force us to structure information in rows and columns will be outdated. Looking at the trend to date, this is already happening. This is supported by [15], who argue that much data today is not natively in structured format; for example, tweets and blogs are weakly structured pieces of text, while images and video are structured for storage and display, but not for semantic content and search: transforming such content into a structured format for later analysis is a major challenge. Thus, he argued that future systems must be able to integrate data and this data will have a more direct linkage with action. The value of data explodes when it can be linked with other data, thus **data integration** is a major creator of value.

[13]Identifies two BI and analytics capabilities: their first argument is that, it enables for business user **data mashup** and modelling; internal platform integration; metadata management; cloud deployment; development and integration and BI platform administration. Data mashup refers to the integration of data from multiple disparate sources to provide users with at-a-glance view of their business performance. This user-driven combination of data from different sources enables the creation of analytic models such as user-defined measures, sets, groups and hierarchies. The "advanced capabilities include semantic auto discovery, intelligent joins, intelligent profiling, hierarchy generation, data lineage and data blending on varied data sources, including multi-structured data".

Of importance is the capability of data mashup in illuminating dark data is the fact that most businesses today have too much data stored in several disparate systems; therefore, unless businesses are able to integrate this data into a single source, then it becomes increasingly difficult to process data that cross different system boundaries which could otherwise have been possible if the data was integrated. Once this data is integrated, analytical tools can then be applied to enable the statistical evaluation of this rich integrated data source and the identification of patterns within the data [25].

[10]defines Analytics as a process that involves the use of statistical techniques (measures of central tendency, graphs, and so on), information system software (data mining, sorting routines), and operations research methodologies (linear programming) to explore, visualize, discover and communicate patterns or trends in data. Simply put, analytics convert data into useful information.

Three distinct types of analytics are explored: *Descriptive, Predictive and Prescriptive*. ***Descriptive analytics*** refers to the application of simple statistical techniques that describes what is contained in a data set or database. ***Predictive analytics*** is the application of advanced statistical, information software, or operations research methods to identify predictive variables and build predictive models to identify trends and relationships not readily observed in a descriptive analysis. Thus, predictive analytics can be applied in the mining of past data, identify patterns and trends on that data and use this new found insight in making future business decisions. For example, multiple regressions can be used to show the relationship between age, weight, and exercise on diet food sales.

Knowing that relationships exist helps explain why one set of independent variables influences dependent variables such as business performance. This capability is important as it helps in building predictive models designed to identify and predict future trends.***Prescriptive analytics*** refers to the application of decision science, management science, and operations research methodologies (applied mathematical techniques) to make best use of allocable resources. The purpose of prescriptive analytics is to allocate resources optimally therefore taking advantage of predicted trends or future opportunities [10].

Business analytics process

The idea of the business process was proposed by [10]; they propose that BA process is premised on answering the question: What valuable or problem-solving information is locked up in the sources of data that an organization has available? Additional questions need to be answered at each of the three steps that make up the BA process. Answering all these questions requires mining the information out of the data via the three steps of analysis that make up the business analytics process.Descriptive statistics is the first stem and it is essential because generally, the size of some data sources can be unimaginable, complex and confusing thus this first stage helps in sorting out the data and making some sense out of its informational value. This may include sorting out the data, cross tabulations and applying the measures of central tendency onto it to restrict the data into some more manageable size. Other measures such as plots, charts and graphs can be applied in order to help decision makers visualize the data in order to understand content opportunities. From this stage, some patterns or business behavior can be identified representing targets of business opportunities and possible future trend behavior. All this can be carried out using an application such as Microsoft excel. This is explained in Figure 2.

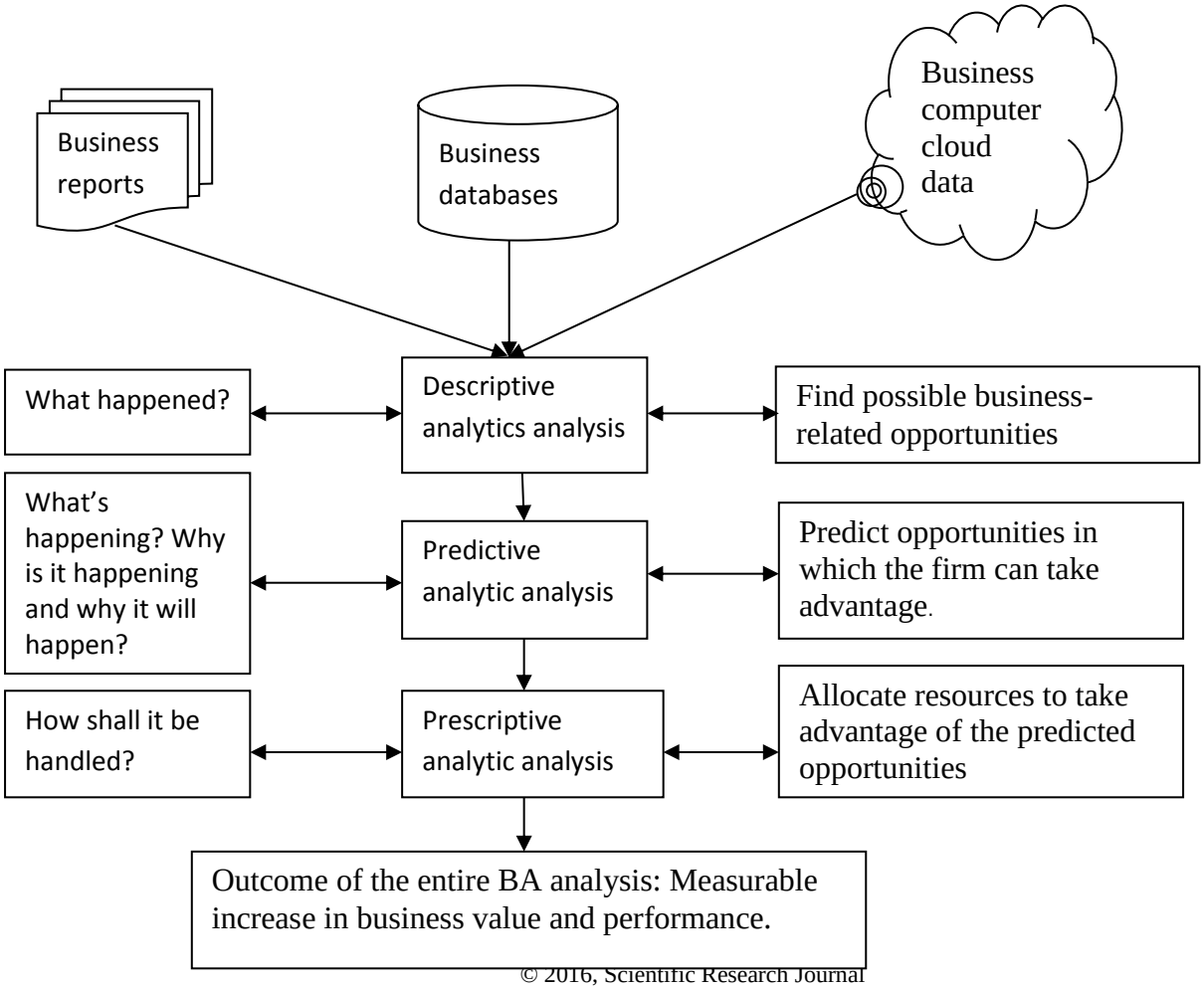

Figure 3: business analytics process. Adopted from **[10]**

Step 2 involves the predictive analytic analysis. This step is done to determine whether the observed trends and behavior found in the analysis of relationship exists or holds true and whether they can be used to predict or forecast the future. Multiple regression analysis is usually applied in this area. This method is ideal to establish whether a statistical relationship exists between the predictive variables found in the descriptiveanalysis. When regression is not practical, other forecasting methods such as exponential smoothing and smoothing averages can be applied as predictive analytics to develop needed forecasts of business trends. Identification of future trends is the main output/activity of this step.

When a firm is able to predict what will happen in the future, it can then be able to take advantage of that and invest appropriately to meet that future need or take advantage of the identified opportunity. Therefore, step 3 (Prescriptive analytics) applies various operations research methodologies to optimally allocate a firm's limited resources to take best advantage of the opportunities it found in the predicted future trends. Limits on human, technology, and financial resources prevent any firm from going after all opportunities they may have available at any one time [10]. Using prescriptive analytics allows the firm to allocate limited resources to optimally achieve objectives as fully as possible. Linear programming for example has been for long applied in the optimization of the supply chain. This stepanswers the question of how best to allocate the resources and manage decision making in the future.

III. Conclusion

Not much research from the academia has been done on the topic of dark data; however, research on big data that exists from the points out the problems that arise because of the continuous accumulation of data by organizations. The practitioner community seems to be way ahead in this area, this is a positive development considering that the end result of any research efforts by the academia is always to assist the practitioner and industry by providing new ideas or supporting existing ideas in the ever dynamic business environment. Challenges of dark data seem to be a key concern for the industry; even thou, research seem to agree on the benefits of tapping onto the vast data from the various sources both internal and external in enhancing business decisions. With the right tools, organizations can now be able to access a large pool of valuable data that provide market trends, customer behaviour and patterns and new lines of investments that can be critical in maintaining a competitive advantage. However, the availability of this data is not a guarantee of its usefulness, investments in data analysis tools that can be able to analyze data that exists in different forms to provide the necessary knowledge and decisions is critical. With the growing adoption of ICT by firms, the concept of dark data will continue to be magnified. This is because as predicted earlier, data will continue to come into organizations in various forms; the embedding of computers into various appliances (ubiquitous computing), will even magnify the challenge of big data further. And the more unstructured data continues to be gathered by firms, the greater the challenge of losing valuable information that can provide essential insights into a firm's activities. Even more, as we move further away from relational databases to object oriented and other types, the more challenging data processing gets. But with the continued availability of data analytics tools, then firms should look forward to better days ahead. There is also need to embed data analytics tools within enterprise systems: this will enable enterprises to have a single analytical repository provides not just the capability to store data but also an integrated environment that enables firms to gather data and analyze it within a single environment.

References

[1] Intel, "Big data policy position paper," 2012.

[2] domo.com, "From big data to better decisions," 2014.

[3] Shane Ryan, "Illuminating Dark Data," 2014.

[4] Wesley M. Cohen and Daniel A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, pp. 128-152, 1990.

[5] Rachel Griffith, Stephen Redding, and John Van Reenen, "R&D and Absorptive Capacity: Theory and Empirical Evidence," *Scand. J. of Economics*, pp. 99-118, 2003.

[6] Tobias Schmidt, "Absorptive Capacity- One size fits all? A Firm-level analysis of absorptive capacity for different kinds of knowledge," in *39th Annual Meeting of the Canadian Economics Association (2005)*, Mannheim,Germany, 2005, pp. 1-37.

[7] Andrew A. King and Karim R. Lakhani, "The Contingent Effect of Absorptive capacity: An Open Innovation Analysis," *Harvard Business School*, pp. 1-35, 2011.

[8] MCKINSEY, "Big Data: The Next Frontier For Innovation, Competition and Creativity," MCKINSEY Global Institute, New York, 2011.

[9] Paul Zikopoulos, Chris Eaton, Diroos Dirk, Tom Deutsch, and George Lapis, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, 1st ed., Steve Sit, Ed. New York, United States: MCGRAW-HILL, 2012.

[10] Marc J. Schniederjans, Dara G. Schniederjans, and Christopher M. Starkey, *Business Analytics Principles, Concepts and Applications*, 1st ed., Gill Editorial Services, Ed. New Jersey, Upper Saddle River: Pearson Education, Inc., 2014.

[11] Viewpointe, "dark data, big data, your data: Creating an action plan for information governance," United states, 2013.

[12] Oracle, "Oracle: Big Data for the Enterprise," Oracle, Redwood, White Paper 2013.

[13] Gartner, "Magic Quadrant for Business Intelligence and Analytic platforms," 2015.

[14] Commvault, "5 ways to illuminate your dark data," 2014.

[15] Divyakant Agrawal, Philip Bernstein, Elisa Bertino, Susan Davidson, and Umeshwar Dayal, "Challenges and Opportunities with Big Data," United States, White Paper 2012.

[16] Kush R. Varshney, "Introduction to Business Analytics," 2012.

[17] ComputerWorld, "Defining business analytics and its impact on organizational decision making," United states, 2009.

[18] Prashant Pant, "Business intelligence (BI) How to build successful BI strategy," Delloite, Practitioner 2009.

[19] IBM, "Business analytics for banking: 3 ways to win," Ottawa, 2010.

[20] Butler analytics, "Business analytics yearbook 2015," 2014.

[21] Marketa Horakova and Hana Skalska, "BUSINESS INTELLIGENCE AND IMPLEMENTATION IN A SMALL ENTERPRISE," *Journal of Systems Integration*, pp. 50-62, 2013.

[22] Gartner, "A step-by-Step Approach to Successful Business Intelligence," 2011.

[23] Nelson Greg, "Introduction to the SAS® 9 Business Intelligence Platform: A Tutorial," in *SESUG proceeding*.

[24] Gregory S. Nelson, "Business Intelligence 2.0: Are we there yet?," *SAS global forum*, pp. 1-10, 2010.

[25] Tanya Elias, "Learning Analytics: Definitions, Processes and Potential," *Creative Commons Attributes: non commercial*, pp. 1-23, 2011.

[26] Oliver E. Williamson, "The Economies of Organization: The Transaction Cost Approach," *American Journal of Sociology*, pp. 548-577, 1981.

[27] Rodrigo Martins, Fernando Ribeiro Serra, Andre da Silva Leite, Manuel Portugal Ferreira, and Dan Li, "Transactions Cost Theory influence in strategy research: A review through a bibliometric study in leading journals," *globADVANTAGE – Center of Research in International Business & Strategy*, pp. 1-29, 2010.

[28] Jongwook Kima and Joseph T. Mahoney, "Property Rights Theory, Transaction Costs Theory, and Agency Theory: An Organizational Economics Approach to trategic Management," *MANAGERIAL AND DECISION ECONOMICS*, pp. 223-242, 2005.