A Study on Challenges and Best Practices of Big Data for SMEs (Small and Medium Enterprises)

Jitendra Gandhi[1], Dr. Pankaj Gandhi[2]
[1]*Himalayan University, Itanagar, Arunachal Pradesh.*
[2]*Associate Professor, P.P.Savani University, Surat, Gujarat.*

Abstract—The term, Big Data' has been coined to refer to the huge bulk of data that cannot be dealt with by traditional data-handling techniques. Big Data is still a new concept, and in the following literature we intend to elaborate it in a palpable approach. Data is a highly valued asset in today's connected world and is growing in volume like never before. Enterprises across the spectrum, from multinationals to small and medium enterprises (SMEs), are exploring avenues to harness and exploit data. The present study has been undertaken to describe the challenges and best practices of big data for SMEs (Small and Medium Enterprises).

Keywords—*Big data, Small and medium enterprises, SMEs*

I.   INTRODUCTION

Big data and business analytics is one of the hottest topics in data analytics and enterprise information systems these days. Data is easier to capture and access through third parties such as Facebook, D&B, and others. There are different types of data which can get from different sources like, geo location data, user-generated content, user's biographic information, machine logging data, and sensor-generated data. Nowadays IT companies find increasing value in leveraging these data to enhance existing applications and create new once made from it. The use of the data is rapidly changing the nature of communication, shopping, advertising, entertainment, and relationship management.

The value of data explodes when it can be linked with other data, thus data integration is a major creator of value. Since most data is directly generated in digital format today, we have the opportunity and the challenge both to influence the creation to facilitate later linkage and to automatically link previously created data. Data analysis, organization, retrieval, and modelling are other foundational challenges. Data analysis is a clear bottleneck in many applications, both due to lack of scalability of the underlying algorithms and due to the complexity of the data that needs to be analyzed. Finally, presentation of the results and its interpretation by non-technical domain experts is crucial to extracting actionable knowledge.

The use of big data technologies is altering the way businesses across industries operate. To address their voluminous data challenges, there is a dire need for SMEs to seriously think about big data adoption.

The above statement is validated by a report published by Research and Markets, which forecasts that big data deployment by SMEs will witness a CAGR of 43% before the year 2018. It further points out that SMEs will continue the momentum of investing in big data and business analytics.

II.   OBJECTIVES

- To understand the benefits of big data for SMEs
- To study the challenges of big data for SMEs
- To study the best practices of big data for SMEs

III.   RESEARCH DESIGN

The researcher has used only secondary data that has been collected from various articles, journals, books, websites etc. It has been used to study the evaluation, conceptual framework, definition, present trends, future prospectus, opportunities & challenges of big data. The researcher also used quantitative research that is the systematic empirical investigation of variables phenomena via statistical & mathematical, theories pertaining to phenomena all the data included is the secondary base & proper references have been given wherever necessary.

IV.   DEFINITION

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. Big data is a collection of data from traditional and digital sources inside and outside that represents a source for ongoing discovery and analysis. Data is easier to capture and access through web sources, data houses, medias or other third parties such as Facebook, D&B, and so on.

An enormous amount of data is created worldwide every day. Much of this data is directly or indirectly relevant for policy and decision-making in the various phases. Big data is big assets for large companies as well as small and medium companies in all sectors as they are making significant advances in their customer relations, product selection and development and consequent profitability through using this valuable commodity. A study has showed that Small and medium enterprises (SMEs) have proved themselves to be slow adopters of the new technology of big data analytics and there may be a danger of being left behind. In India, SMEs are a vital part of the economy, and the challenges they encounter need to be addressed as a matter of urgency. This research paper identifies barriers to SME in adoption of big data analytics and recognizes their complex challenge to all

stakeholders, including national and international policy makers, IT, business management and data science communities. This paper also emphasizes the best practices and benefits of big data in small and medium enterprises.

"Big data" is literally just a lot of data. While it's more of a marketing term than anything, the implication is usually that you have so much data that you can't analyze all of the data at once because the amount of memory (RAM) it would take to hold the data in memory to process and analyze it is greater than the amount of available memory. This means that analyses usually have to be done on random segments of data, which allows models to be built to compare against other parts of the data. To break that down in simple words, let's say that Facebook wants to know which ads work best for people with college degrees. Let's say there are 200,000,000 Facebook users with college degrees, and they have been each served 100 ads. Now the ads which posted to these people contain more than 1000s of features such as color, design, picture, presence of male or female, presence of any area, size of ad, purpose of ad etc. and so on. There are different segment people for each feature. Big data helps you to analyze each feature with their specific segment and the pattern of the users. Like through big data analysis we can know most prominent color that college people like the most or say presence of male or female picture which attract the most in the ads.

The goal is to just figure out how which features are most effective in getting college grads to click ads. Maybe you get model on a random sample of 1,000,000 users finds that ads with people or with cute children or about food get the most clicks. From this model, you can predict that what college grads exactly want, and you can then test that to see how well your prediction holds up.

We can also select another subset by getting people who likes the most ads which shows family picture and get more clicks. Thus, with the help of big data we can get the clear picture of the pattern of consumer taste and preference and accordingly can plan the business.

The use of big data technologies is altering the way businesses across industries operate. To address their voluminous data challenges, there is a dire need for SMEs to seriously think about big data adoption.

The above statement is validated by a report published by Research and Markets, which forecasts that big data deployment by SMEs will witness a CAGR of 43% before the year 2018. It further points out that SMEs will continue the momentum of investing in big data and business analytics.

V. CATEGORIES OF BIG DATA

Big data' has four dimensions, often addressed by the acronym VVVC1:
high-volume (V) data,
created with high velocity (V)
in great variety (V) and

of high complexity (C).

Regarding volume, there are lots of data are created worldwide per day and it is expected that the creation volume is doubling every 40 months. But the thing is the storage of the data and analysis of the relevant data which done by Big data. Lots data doesn't mean lots of asset rather important data or say nicely analyzed data are the asset of the organization.

Regarding velocity, modern information technology (IT) infrastructures enable data to be submitted for analysis in nearly real time.

Regarding variety, the various data sources bring about data of different formats, like classical database formats, textual data, image data, sensor data, structured, semi-structured and totally unstructured data.

Regarding complexity, important aspects of data complexity are multivariate, multiformat, multidate and multiresolution. Many covariate data for one target entity, for example, a sales process, arise from different sources in different formats at different rates of acquisition and granularity.

Variety of data includes following formats of data:

A. **Structured**

B. **Unstructured**

C. **Semi-structured**

A. *Structured*

Any data that can be stored, accessed and processed in the form of predefined format is termed as a 'structured' data. Over the period, capacity in computer science have achieved greater success in developing techniques for working with such kind of data (where the format is well known in advance) and deriving value out of it.

Examples of Structured Data

An 'Employee' table in a database is an example of Structured Data

Emp_ID	Employee_Name	Gender	Department	Age	Salary_In_lacs (Rs.) P.A.
MIPL 0101	Ved Padaria	Male	Finance	30	650000
MIPL 0201	Pari Patel	Female	Admin	25	650000
MIPL 0202	Shaurya Gandhi	Male	Human Resource	28	500000

B. Unstructured

Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it. Typical example of unstructured data is, a heterogeneous data source containing a combination of simple text files, images, videos etc. Now a day organization have wealth of data available with them but unfortunately, they don't know how to derive value out of it since this data is in its raw form or unstructured format.

C. Semi-structured

Semi-structured data can contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS. Example of semi-structured data is a data represented in XML file.

Examples of Semi-Structured Data

Personal data stored in a XML file

<rec><name>Ved Padaria</name><sex>Male</sex><age>30</age></rec>
<rec><name>Pari Patel</name><sex>Female</sex><age>25</age></rec>
<rec><name>Shaurya Gandhi</name><sex>Male</sex><age>28</age></rec>

VI. WHAT IS UNIQUE ABOUT BIG DATA?

Companies have sought for decades to make the best use of information to improve their business capabilities. However, it's the structure and size of Big Data that makes it so unique. Big Data is also special because it represents both significant information - which can open new doors - and the way this information is analyzed to help open those doors. The analysis goes hand-in-hand with the information, so in this sense "Big Data" represents a noun - "the data" - and a verb - "combing the data to find value."

VII. BENEFITS OF BIG DATA FOR SMEs

Businesses can utilize outside intelligence while taking decisions

Access to social data from search engines and sites like Facebook, twitter is enabling organizations to fine tune their business strategies.

Improved customer service

Traditional customer feedback systems are getting replaced by new systems designed with 'Big Data' technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.

Better operational efficiency

'Big Data' technologies can be used for creating staging area or landing zone for new data before identifying what data should be moved to the data warehouse. In addition, such integration of 'Big Data' technologies and data warehouse helps organization to offload infrequently accessed data.

Early identification of risk to the product/services, if any

Because of having patterns through big data, SME's can assume risk in advance and prepare to take corrective measures for the same.

VIII. CHALLENGES OF BIG DATA FOR SMEs

A. Storage cost is high

Tones of data is generated in Big Data and the SMEs can only maintain and manage the storage cost of handling this Big Date if they have sufficient knowledge about it. They can opt to choose a hybrid cloud computing model for data security reasons and justify the expenses with this knowledge only.

B. Doubt regarding spend on time and energy

The doubt regarding spend can only be cleared ones the extracted data proves its value by opening gateways to growth and challenge larger organizations. It is hard to realize the potential of Big Data without exploring and exploiting it within limited capacity of SMEs.

C. Insufficient data expertise

SMEs lack knowledge about the useful data mining tools which are practical and are based on affordable technology. With this knowledge, they can benefit themselves in processing Big Data and apply it pragmatically.

D. Data collection sources are ambiguous

A lot of data is generated by social media, emails CRM applications etc. SMEs can start with focus on smaller data sets for these data sources and try to understand the behaviour patterns of the customer. The company's data serves the purpose to a great extent, but when engaged with channel partners, the data can be aggregated and managed for best results. This data mining is difficult for the company to manage on own and this way the expertise of channel partners play a vital role in understanding and using the Big Data.

E. Understanding and prioritizing the data from garbage that is coming into the enterprise

F. Retention of the existing data which is sometimes relevant for smaller duration and sometimes relevant after a longer duration and in addition the new Data gets added very rapidly making it an expensive affair to store.

G. Getting skills of Big Data solutions is challenging with the advent of new technologies and tools.

H. Promptness in analysing a larger data which takes longer time is required as in most of the situations the results of the analysis are required for immediate use.

I. Incomplete data hinders the process of data analysis and completing and correcting this is a challenge. Managing this probabilistic data is the need of the time.

J. Big Data has huge privacy concerns attached to it.

IX. BEST PRACTICES OF BIG DATA FOR SMEs

A. Big Data to be aligned to get specific Business Goals

Finding valuable data from the extensive data, which can be used for attaining the business goal, must be smart work. For this smart work need new investments in skills, organization, and infrastructure with the common goal to excel in all stages of business. Example – Understanding the purchasing patterns of the buyers by analyzing the data related to their search online for product or service and aligning it with the production to meet the demands at the specific time of the year. In short, making the business plans as per the needs of the buyers to facilitate best of the services for attaining the business goal of profitability. If not done properly, the data will drive the business haywire.

B. Ease Skills Shortage with Standards and Governance

IT governance program with rules of standard technology and necessary skill set requirements should be added in Big Data technologies, considerations and decisions which will allow manage costs and best leverage of the present resources to make things happen an authentic way. The SMEs opting to implement Big Data Solutions and strategies should possess the necessary skills requirement, if not it should overcome the shortcoming by training the existing resources, hiring new resources or leveraging with consulting firms.

C. Sharing of knowledge culture should be promoted

Bid Data being a new and expanding investment, it is advisable for the SMEs to share solution knowledge related topic with others across the enterprise. SMEs can post their questions and queries and share information on common groups like in LinkedIn so that people are benefited on a continuous basis and time is not wasted on finding same solutions again and again by different organizations. This approach can help getting faster results from the findings of Big Data.

D. Prioritize to convert Unstructured to structure data

In analyzing Big Data, for business clarity is better by connecting and integrating low density Big Data with the structured data. If any organization is analyzing Data for any sentiment, it will furnish with better clarity and conclusions if all variety of customers from varied criteria are analyzed rather than analyzing a small group of best customer's nature. Data related to customer, product, equipment and environmental will add more value if they are added to more relevant data points to core master and analytical summaries.

E. Plan work and monitor as per plan

For perfect planning, the management and IT need to bring clarity to the direction and requirement from the findings of the Big Data. So, there is need for the Analyst and Data scientist to work closely with the business, understand key business knowledge gaps and requirements. In order to avail the interactive exploration of data and the experimentation of such statistical algorithms, SMEs need to set up high performance work areas which can monitor that the Data is put to use as per the plan and things are going in the planned direction to achieve the company goals.

F. Align with Cloud Operating Model

Huge resources are required for Big Data processes. Both interactive experimentation and running production jobs require resources for transactions, master data, reference and summarization of the Big Data. For this analytical sand boxes should be created to experiment and ensure control of the entire data flow, pre-processing, integration, in-database summarization, post-processing and analytical modelling. The decision for resorting for public or private cloud provision and security strategy also plays a vital role in supporting this requirement of Big Data analysis.

G. Evaluate the Big Data needs

The reasons for resorting to Big Data should be identified, whether it is for studying customer behavior, improve product design, add new product feature or to add a new product, or implementation of some process improvement. Just collecting all data will add load to the present data load and make the process difficult. In short, the specific need has to be identified so that resources are not wasted.

H. Analyze data within budgetary constrains

It is advised to use existing tools and techniques to analyze data. Try this with a smaller data and evaluate the finding if they match your expectation or not. In case they don't match, the approach needs to be redefined. If they match with expectation, same approach can be used for bigger scale and bigger data. These initial tests will ensure that the organization is not deviating for the goals and no resources are waited. If all the results and findings are of use, then only SMEs should

invest in big storage serves to get assured ROI.

I. Subscribe to cloud storage services

Ones the SME is sure about resorting to Big Data to achieve the business goals, they can decide for the storage services. It is advised to opt for cloud storage service which can be scaled up or down as per the business needs. Cloud services has the facility to create own analytic sandbox and keep expensed in control by safeguarding the capital investment.

X. CONCLUSION

It is the right time for SMEs to harness the power of Big Data Analytics. Their focus should be to find the right solution. This will enable them to develop a unique image in the market and gain a competitive edge.

Affordability is a major concern for SMEs while delving into Big Data Analytics. However, with proper control of costs, execution, and resources; SMEs can surmount their fears.

They should focus on key problems and define quantitative methods to understand problems. Likewise, they can create a lot of value for their customers.

SMEs can hire external firms or interns to conduct big data analytics. However, it is vital to have a clearly defined question before you move forward. With the clarity of goal in mind, hiring the right person becomes a lot easier.

They should focus on key problems and define quantitative methods to understand problems. Likewise, they can create a lot of value for their customers. SMEs can hire external firms or interns to conduct big data analytics. However, With the clarity of goal in mind, hiring the right person becomes a lot easier.

XI. REFERENCES

[1]. About big data (4th April 2017) Retrieved from http://www.zettaset.com/index.php/info-center/what-is-big-data/

[2]. What is big data? (4th April 2017) retrieved from https://www.oracle.com/big-data/index.html

[3]. https://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/

[4]. Big data for SMEs (19th May 2017) retrieved from https://www.indusnet.co.in/can-smes-afford-big-data-analytics/

[5]. http://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf

[6]. http://ijcsmc.com/docs/papers/July2014/V3I7KJ06.pdf http://clcloud.com/DYNAA2014/papers/P001DYNAA2014.pdf

[7]. https://www.ijarcce.com/upload/2016/february-16/IJARCCE%2015.pdf

[8]. https://wikipedia.org/wiki/Big_data

[9]. What is use of big data (20th May 2017) retrieved form https://www.quora.com/Big-Data-What-is-big-data-and-what-is-its-use

[10].categories of big data (20th May 2017) retrieved from http://www.techrepublic.com/blog/big-data-analytics/big-data-basic-concepts-and-benefits-explained/

Mr. Jitendra Gandhi, is a Ph.D scholar of Himalayan University. He is pursuing investigation on "A Study on the effectiveness of E-commerce Application in Small & Medium Sized Enterprise In Gujarat State" Under the guidance of Dr. Pankaj Gandhi.

Dr. Pankaj J Gandhi is an Associate Professor, P.P.Savani University, Surat, Gujarat, India. He has done doctorate in the field of HRM & Engineering. He has published 7 research paper in international journals till date. He has done MBA from IIM (Kerala). He is associated with G. N. Pharma – Medico Research as a Chief Scientific Officer Since August 2011.