Big data security and privacy issues in SMEs

Reena Singh

Department of CSE, Suresh Gyan Vihar University Jaipur Raj. India

Abstract: In all economies, especially in developing and transition economies, there is now a consensus among state policy makers, development economists as well as international development partners that small and medium enterprises (SMEs) are a potent driving force for their industrial growth and indeed, overall economic development. In recent times, the concept of Big Data has been seen as a new solution to help in policy and practice in all sorts of application context and domains. The impact of abundance data collected and stored over a number of years by various organisations both public and private has led to many innovative data analytics technologies. The thrust of this paper therefore is focusing on SME growth, that is, how to assist regional small business growth using Big Data. Harnessing big data practice for SME growth has potential to challenge current decision making and policy initiatives both at the government level (macro), as well as at the SME level (micro). Thispaper will assess the extent to which Big Data can be harnessed for SME growth; and develop a systems based method for making intervention based on Big Data practice for SME growth.

Keywords: SMEs; Big data ; Technology in SMEs ; Security ; Privacy ; Analytics

I. INTRODUCTION

In present scenario it is very common thinking that Information Systems are useful for the organisations in gaining the competitive advantage over the others, it helps managers to take decisions according to the requirement of the situation and according to the available resources. Indian SMEs are considered as the backbone of economy contributing to 45% of the industrial output, 40% of India's exports, employing 60 million people, create 1.3 million jobs every year and produce more than 8000 quality products for the Indian and international markets. With approximately 30 million SMEs in India, 12 million people expected to join the workforce in next 3 years and the sector growing at a rate of 8% per year, the SMEs are deploying information technology to take the substantial advantage from it.

The SMEs in India facing various challenges such as the absence of adequate and timely institutional credit facilities, limited capital and knowledge, lack of access to technology and skilled manpower, competition from large enterprises and globalisation. These issues need to be addressed to tap the full potential of the sector, which brings about social and economic development of the country. SMEs are facing competition from multinational corporations in the domestic market [1]. Small & Medium Enterprises Development Act, 2006 the Small and Medium Enterprises (SME) are classified in two Classes respectively.

1. *Manufacturing Enterprises:* The enterprises engaged in the manufacture or production of goods pertaining to any industry specified in the first schedule to the industries (Development and regulation) Act, 1951). The Manufacturing Enterprise is defined in terms of investment in Plant & Machinery.

2. *Service Enterprises:* The enterprises engaged in providing or rendering of services and are defined in terms of investment in equipment.

Table I: Definition of SME as per SME Act, 2006

Manufacturing Sector	
Enterprises	Investment in plant and machinery
Small Enterprises	More than twenty five lakh rupees but does not exceed five crore rupees
Medium Enterprises	More than five crore rupees but does not exceed ten crore rupees
Service Sector	
Enterprises	Investment in equipment's
Small Enterprises	More than ten lakh rupees but does not exceed two crore rupees
Medium Enterprises	More than two crore rupees but does not exceed five core rupees

A. Importance of Big Data

The government's emphasis is on how big data creates "value" –both within and across disciplines and domains. Value arises from the ability to analyze the data to develop actionable information. There cab be five generic ways that big data can support value creation for organizations.

1. Creating transparency by making big data openly available for business and functional analysis (quality, lower costs, reduce time to market, etc.)

2. Supporting experimental analysis in individual locations that can test decisions or approaches, such as specific market programs.

3. Assisting, based on customer information, in defining market segmentation at more narrow levels.

4. Supporting Real-time analysis and decisions based on sophisticated analytics applied to data sets from customers and embedded sensors.

5. Facilitating computer-assisted innovation in products based on embedded product sensors indicating customer responses.

B. Big data Characteristics

One view, espoused by Gartner's Doug Laney describes Big Data as having three dimensions: volume, variety, and velocity. Thus, IDC (International Data Corporation) defined it: Big data technologies describe 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." Two other characteristics seem relevant: value and complexity. We summarize these characteristics as given below.

1. Data Volume:

Data volume measures the amount of data available to an organization, which does not necessarily have to own all of it as long as it can access it. As data volume increases, the value of different data records will decrease in proportion to age, type, richness, and quantity among other factors.

2. Data Velocity:

Data velocity measures the speed of data creation, streaming, and aggregation. Ecommerce has rapidly increased the speed and richness of data used for different business transactions (for example, web-site clicks). Data Variety: Data variety is a measure of the richness of the data representation – text, images video, audio, etc.

3. Data Value:

Data value measures the usefulness of data in making decisions. It has been noted that "the purpose of computing is insight, not numbers". Data science is exploratory and useful in getting to know the data, but "analytic science" encompasses the predictive power of big data.

4. Complexity:

Complexity measures the degree of interconnectedness (possibly very large) and interdependence in big data structures such that a small change (or combination of small

changes) in one or a few elements can yield very large changes or a small change that ripple across or cascade through the system and substantially affect its behavior, or no change at all.

In addition to big data challenges induced by traditional data generation, consumption, and analytics at a much larger scale, newly emerged characteristics of big data has shown important trends on mobility of data, faster data access and consumption, as well as ecosystem capabilities.

In this paper, We studied a system that can scale to handle a large number of sites and also be able to process large and massive amounts of data. However, state of the art systems utilizing HDFS and Map Reduce are not quite enough/sufficient because of the fact that they do not provide required security measures to protect sensitive data. Moreover, Hadoop framework is used to solve problems and manage data conveniently by using different techniques.

C. Types of Big Data and Sources

There are two types of big data: structured and unstructured.

1. Structured Data:

Structured Data are numbers and words that can be easily categorized and analyzed. These data are generated by things like network sensors embedded in electronic devices, smart phones, and global positioning system (GPS) devices. Structured data also include things like sales figures, account balances, and transaction data.

2. Unstructured Data:

Unstructured Data include more complex information, such as customer reviews from commercial websites, photos and other multimedia, and comments on social networking sites. These data cannot easily be separated into categories or analyzed numerically. The explosive growth of the Internet in recent years means that the variety and amount of big data continue to grow. Much of that growth comes from unstructured data.

Fig. 1: Sources of Big Data

II. BIG DATA CHALLENGES TO INFORMATION SECURITY AND PRIVACY

With the proliferation of devices connected to the Internet and connected to each other, the volume of data collected, stored, and processed is increasing everyday, which also brings new challenges in terms of the information security. In fact, the currently used security mechanisms such as firewalls and DMZs cannot be used in the Big Data infrastructure because the security mechanisms should be stretched out of the perimeter of the organization's network to fulfill the user/data mobility requirements and the policies of BYOD (Bring Your Own Device). Considering these new scenarios, the pertinent question is what security and privacy policies and technologies are more adequate to fulfill the current top Big Data privacy and security demands (Cloud Security Alliance, 2013). These challenges may be organized into four Big Data aspects such as infrastructure security (e.g. secure distributed computations using MapReduce), data privacy (e.g. data mining that preserves privacy/granular access), data management (e.g. secure data provenance and storage) and, integrity and reactive security (e.g. real time monitoring of anomalies and attacks).

Considering Big Data there is a set of risk areas that need to be considered. These include the information lifecycle (provenance, ownership and classification of data), the data creation and collection process, and the lack of security procedures. Ultimately, the Big Data security objectives are

no different from any other data types – to preserve its confidentiality, integrity and availability.

Being Big Data such an important and complex topic, it is almost natural that immense security and privacy challenges will arise (Michael & Miller, 2013; Tankard, 2012). Big Data has specific characteristics that affect information security: variety, volume, velocity, value, variability, and veracity (Figure 1). These challenges have a direct impact on the design of security solutions that are required to tackle all these characteristics and requirements (Demchenko, Ngo, Laat, Membrey, & Gordijenko, 2014). Currently, such out of the box security solution does not exist.

Fig. 2: The five V's of Big Data

Cloud Secure Alliance (CSA), a non-profit organization with a mission to promote the use of best practices for providing security assurance within Cloud Computing, has created a Big Data Working Group that has focused on the major challenges to implement secure Big Data services (Cloud Security Alliance, 2013). CSA has categorized the different security and privacy challenges into four different aspects of the Big Data ecosystem. These aspects are Infrastructure Security, Data Privacy, Data Management and, Integrity and Reactive Security. Each of these aspects faces the following security challenges, according to CSA:

1. *Infrastructure Security :*
   1. Secure Distributed Processing of Data
   2. Security Best Actions for Non-Relational Data-Bases
2. *Data Privacy :*
   3. Data Analysis through Data Mining Preserving Data Privacy
   4. Cryptographic Solutions for Data Security
   5. Granular Access Control
3. *Data Management and Integrity :*
   6. Secure Data Storage and Transaction Logs
   7. Granular Audits
   8. Data Provenance
4. *Reactive Security :*
   9. End-to-End Filtering & Validation
   10. Supervising the Security Level in Real-Time

These security and privacy challenges cover the entire spectrum of the Big Data lifecycle (Figure 2): sources of data production (devices), the data itself, data processing, data storage, data transport and data usage on different devices.

Fig. 3: Security and Privacy challenges in Big Data
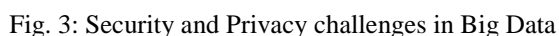
It is clear that Big Data present interesting opportunities for users and businesses; however these opportunities are countered by enormous challenges in terms of privacy and security (Cloud Security Alliance, 2013). Traditional security mechanisms are insufficient to provide a capable answer to those challenges.

III. BIG DATA PROBLEMS AND CHALLENGES

The problem comes straight way when the data tsunami requires us to make specific decisions, about what data to keep and what to reject, and how to store what we keep reliably with the right metadata. Transforming unstructured content into structured format for later analysis is a major challenge. Data analysis, organization, retrieval, and modelling are other foundational challenges. Since most data is directly generated in digital format today, the challenge is to influence the creation so as to facilitate later linkage and to automatically link previously created data.
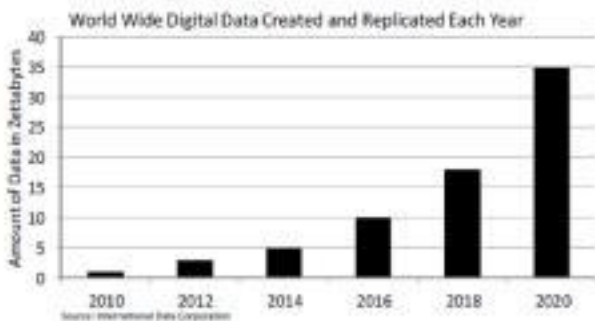

Fig 4: Worldwide Data Creation

1. Heterogeneity:

Machine analysis algorithms expect homogeneous data, and cannot understand fine distinction. Computer systems work most expeditiously if they can store multiple items that are identical in size and structure. So, data must be carefully structured as a first step in (or prior to) data analysis.

2. Scale:

The first thing that anyone thinks of in Big data is its size. Managing large and rapidly increasing volumes of data has been an exigent issue for many decades. In the past, this challenge was palliated by processors getting faster. But there is a fundamental shift happening now: 'data volume is scaling faster than computer resources'. Unluckily, parallel

data processing techniques useful in the past for processing data across nodes don't directly apply for intra-node parallelism, since the architecture is very different.

3. Timeliness:

The larger the data set to be processed, the longer it will take to analyze. The design of a system that effectively deals with size is likely also to result in a system that can process a given size of data set faster. However, it is not just this speed that is usually meant when one speaks of Velocity in the context of Big Data. There are many situations in which the result of the analysis is required immediately. Given a large data set, it is often essential to find elements in it that meet a precise criterion. Scanning the entire data set to find suitable elements is obviously unfeasible.

4. Privacy and Security:

A key value proposition of big data is access to data from multiple and diverse domains, security and privacy will play a very important role in big data research and technology. In domains like social media and health information, more data is gathered about individuals, so there is a fear that certain organizations will know too much about individuals.

Developing algorithms that randomize personal data among a large data set so as to ensure privacy is a key research problem.

5. System Architecture:

Business data is examined and studied for many purposes that might include system log analytics and social media analytics for risk measurement, customer retention, and brand management etc. Typically, such diverse tasks have been handled by separate systems, even if each system includes common steps. The challenge here is not to build a system that is ideally suited for all processing tasks. Instead, the need is for the primary system architecture to be flexible enough that the components built on top of it for showing the various

kinds of processing tasks can tune it to expeditiously run these different workloads.

IV. PROBLEMS OF SMALL AND MEDIUM ENTERPRISES

Baadom (2004) asserted that the following problems militate against the effective operation of small and medium enterprises;

1. *Poor Implementation of Policies:* there have been many good policies formulated in the past by the government in developing countries to improve SMEs, but weak implementation has made it impossible to realize the goal.

2. *Lack of Continuity:* most small scale establishments are sole proprietorship and such establishment often ceases to function as soon as the owner loses interest or dies. This raises the risk of financing such business.

3. *Poor Capital Outlay:* inadequate capital outlay has often affected small scale business adversely. Financiers often regard the sector has high risk area and therefore feel skeptical about committing their fund to it.

4. *Poor Management Expertise:* Management has always been a problem in this sector as most small scale businesses do not have the required management expertise to carry them through once the business start growing. The situation gets compounded as training is not usually accorded priority in such establishments.

5. *Inadequate Information Base***:** Small scale business enterprises are usually characterized by poor record keeping and that usually starve of necessary information required for planning and management purposes. This usually affects the realization of the sector.

6. *Lack of Raw Materials:* In some small scale business enterprises, raw materials are sourced externally, hence the fate of such enterprises to foreign exchange behavior. The fluctuation of foreign exchange may therefore make it difficult to plan and that may precipitate same stock that may destabilize the setup.

7. *Poor Accounting System:* the accounting system of most small scale business enterprises lack standard and does not make room for the assessment of their performances. That creates opportunity for mismanagement, which consequently may lead to enterprise failure.

8. *Unstable Policy Environment:* Government policy instability has not been helpful to small scale businesses. That has been destabilizing and has indeed sent many SMEs to early fold-ups.

REFERENCES

[1] Nabeel Khan and Adil Al-Yasiri. 2015. Framework for cloud computing adoption: a roadmap for smes to cloud migration. School of Computing Science and Engineering, University of Salford, Manchester, United Kingdom..

[2] Nazir Ahmad Research Scholar and Jamshed Siddiqui Associate Professor.2013. Implementation of IT/IS in Indian SMES: Challenges and Opportunities Department of Computer Science Aligarh Muslim University Aligarh, India.

[3] Ogbuokiri, B.O.(MSc.) .2015. Implementing bigdata analytics for small and medium enterprise (SME) regional growth. Department of Computer science, University of Nigeria, Nsukka, Enugu state.

[4] Dr. Jangala. Sasi Kiran .2015. Recent Issues and Challenges on Big Datain Cloud Computing. Dept. of CSE, Vidya Vikas Institute of Technology, Chevella, R.R. Dt –Telangana, INDIA.

[5] Venkata Narasimha Inukollu.2014. Security issues associated with big data incloud computing. Department of Computer Engineering, Texas Tech University, USA.

[6] Priya P. Sharma .2014. Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution. Information Technology Department SGGS IE&T, Nanded, India.

[7] José Moura . Security and Privacy Issues of Big Data. ISCTE-IUL, Instituto Universitário de Lisboa, Portugal .

[8] Mr. Swapnil A. Kale1.Understanding the big data problems and their solutions using hadoop and map-reduce.2014. ME (CSE), First Year, Department of CSE, Prof. Ram Meghe Institute Of Technology and Research, Badnera,Amravati. Sant Gadgebaba Amravati University, Amarvati, Maharashtra, India - 444701.

Author Information

Mrs.Reena Singh received masters degree of MCA from UPTU Lukhnow UP. Her interest areas are Big data, Cloud Computing, and Social Engineering. Currently she is Pursuing Phd in CSE from Gyan Vihar University Jaipur Raj. India. Email : reena_mmm@rediff.com