

# Big Data Expansion and Challenges

Pavan Kumar Vadrevu\*<sup>1</sup>, Ravi Kumar Suggala\*<sup>2</sup>, G Tej Varma\*<sup>3</sup>  
<sup>1,2,3</sup> Assistant Professor, Dept of IT, SVCE,  
 Bhimavaram

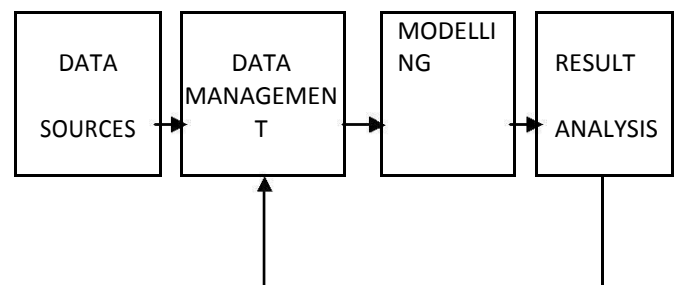
**Abstract:-** Improvement in technology and better affordability of digital device data have presided over today the age of Big Data, parasol term for the explosion in the quantity and diversity of high frequency data. The data hold the impending as largely untouched to allow decision makers to track development progress, improve social protection, and understand where existing policies and developers require adjustment. Call logs, mobile-banking transactions, and online user generated content such as blog posts and Tweets, online searches, satellite images, etc. into actionable information require computational techniques to expose trends and patterns within and between these tremendously large socioeconomic datasets. Fresh insights gleaned from such data mining should complement official statistics, survey data, and information generated by Early Warning Systems, adding depth on human behaviors and experiences and doing in real time, thereby lessening both information and time gaps. With the swear come questions about the analytical value and thus policy significance of data including concerns over the relevance of the data in developing contexts, its representativeness, its reliability as well as the over arching confidentiality issues of utilizing personal data. This paper offer a theory of technology driven social change in the Big Data. It aims to define the main concerns and challenges raised by “Big Data” as concretely and openly as possible, and to suggest ways to address at least a few aspects of each. It is important to distinguish that Big Data and real time analytics are no modern solution for age old development challenges.

## 1. INTRODUCTION

Big data is creating strange opportunities for businesses to achieve deeper, faster insight that can make stronger decision making, improve the customer experience, and accelerate the speed of innovation. But today, most big data yields neither meaning nor value. Businesses are so inundated by the amount and variety of data cascading into and through their operations that they struggle just to store the data much less analyze, interpret, and present it in meaningful ways. For assist businesses are increasingly turning to challenges based data discovery tools leading to estimate a 30 percent compound annual growth rate through 2015. The tools promote self service business intelligence that enable a large number of users to easily integrate data from a wide range of sources like click streams, social media, log files, videos, and more. With the help of high powered desktops and mobile computing devices. Users can perform real time predictive analyses and showcase the results in compelling, interactive, and easily implicit visual formats. The trend toward challenges based data discovery tools is worth exploring by any business that seeks to derive more value from big data. The potential business benefits are immense, and data governance best practices can be used to help ensure a safe transition. As demonstrated by usage real world

applications of challenges based data discovery tools are already delivering greater customer and market insights to businesses around the world.

The Trend toward Challenges based Data Discovery



## 2. DATA CHALLENGES

One of the challenges in analyzing big data relates to its velocity. The rapid generation of big data can lead to significant business insights and predictions, but only if real time data can be analyzed quickly in hours rather than weeks or months. Dropping the latency from data capture to action is completely fundamental. Today, however IT Manager survey found that only about half of IT managers perform data analytics in real time, while the other half continue to rely on batch processing that fails to capture the closeness of big data. A final challenge driving the rise of challenges based data discovery tools is the increasing availability of mobile devices.

Businesses that continue to rely on central creation of reports by a few highly trained experts are missing an opportunity to adopt a faster more cost effective and more democratized business model that takes advantage of the junction of big data and the mobile workforce to speed insights and improve collaboration. Data analytics and challenges are not new. For decades, businesses have collected data, analyzed it using a variety of tools and generated reports.

The process may take weeks or months but eventually a few highly trained data analysts are able to pull the necessary figures from their dashboards and issue static, rearview reports to executives and other employees. business are finding that this traditional reporting process does not work nearly as well for big data and definitely is not sufficient to capture the potential value that big data represents. The primary challenges shoot from what are commonly termed the “3 Vs” of big data: volume, variety, and velocity.

Most conventional reporting and data mining tools cannot handle the vast volume of data although the variety and velocity of the data often present even greater challenges. Big data includes three types of data structured, semi structured, and unstructured

Intel's IT Manager Survey of 200 IT professionals found that four of the top five data sources for IT managers today are semi structured or unstructured. In fact, an IBM survey of more than 1,100 business and IT professionals found that fewer than 26 percent of respondents who had active big data efforts could analyze extremely unstructured data such as voice and video and just 35 percent could analyze streaming data.

Applying Big Data analytics to the fuel of development faces several challenges. Some relate to the data including its attainment and sharing, and the overarching concern over privacy. Others relate to its analysis. This section discusses the most outstanding of the challenges

### 3. DATA CONFIDENTIALITY

Confidentiality is the most sensitive issue with conceptual, legal, and technological implication. In its narrow sense confidentiality is defined by the International Telecommunications Union as the

“Right of individuals to control or influence what information related to them may be disclosed” Confidentiality can also be understood in a broader sense as encompassing that of companies wishing to protect their competitiveness and customers and states excited to preserve their power and society. In both these interpretations, confidentiality is an overarching concern that has a wide range of implications for anyone wishing to explore the use of big data for development.

Data gaining, storage, preservation, use and presentation. Confidentiality is a fundamental person right that has both intrinsic and active values. Two authors, Helbing and Baliotti stress the necessity to ensure an appropriate level of confidentiality for individuals, companies and societies at large. In their words, “a modern society needs confidentiality in order to flourish” Without privacy, safety, variety, pluralism, innovation, our basic freedoms are at risk.

Importantly these risks concern even individuals who have nothing to hide. There is no need to expand at length on the importance and sensitivity of information for corporations and states. Focusing on individual confidentiality it is likely that in many cases the primary producers i.e. the users of services and devices generating data are ignorant that they are doing so and or what it can be used for. For example, people regularly consent to the collection and use of web generated data by simply ticking a box without fully realizing how their data might be used or misused.

It is also uncertain whether bloggers and Twitter users for instance actually consent to their data being analyzed. In

addition fresh research showing that it was possible to de anonymous previously anonymised datasets raises concerns.

The wealth of individual level information that Google, Face book, and a few mobile phone and credit card companies would jointly hold if they ever were to pool their information is in itself concerning. Because confidentiality is a pillar of democracy, we must remain alert to the possibility that it might be compromised by the rise of new because confidentiality is a pillar of democracy, we must remain alert to the possibility that it might be compromised by the rise of new technologies, and put in place all necessary safeguards.

### 4. DATA ADMISSION AND DATA DISTRIBUTION

Even though much of the openly available online data has potential value for development, there is a immense contract more valuable data that is closely held by corporations and is not accessible. One challenge is the unwillingness of private companies and other institutions to share data about their clients and users as well as about their own operations. Obstacles may include legal or reputational considerations, a need to protect their competitiveness, a civilization of secrecy, and, more broadly, the absence of the right incentive and information structures. There are also institutional and technical challenges when data is stored in places and ways that make it difficult to be accessed, transferred etc.

For example, MIT professor Nathan Eagle often anecdotally describes how he spent weeks in the basements of mobile phone companies in Africa searching through hundreds of boxes filled with magnetic backup tapes to gather data. An Indonesian mobile carrier estimated that it would take up to half a day of work to extract one day's worth of backup data currently stored on magnetic tapes. Even within the UN system it can prove difficult to get agencies to share their programmed data for a combination of some or all of reasons listed above.

Engaging with appropriate partners in the public and private sectors to access non public data entails put in place non trivial lawful preparations in order to secure reliable access to data stream and get access to back up data for display analysis and data training purposes. There are other technical challenges of inter comparability of data and inter operability of systems but these might be relatively less problematic to deal with than getting formal access or agreement on licensing issues around data. For Big Data development to gain grip these are serious make or break challenges. Any initiatives in the field have to to fully know the salience of the confidentiality issues and the importance of handling data in ways that ensure that confidentiality is not compromised. These concerns must nurture and shape ongoing debate around data confidentiality in the digital age in a positive manner in order to plan strong principles and strict rules backed by adequate tools and systems to ensure confidentiality preserving analysis.

At the same time the guarantee will not be fulfilled if institutions primarily private corporations decline to share data altogether. In glow of these necessities Global Pulse for instance is putting forth the concept of data philanthropy whereby corporations take the initiative to anonymize their data sets and provide this data to social innovators to mine the data for insights, patterns and trends in real time or near real time. Whether the concept of data philanthropy takes hold or not it certainly points to the challenges and avenues for reflection in the future and we can expect to see further refinements and alternative models proposed for how to deal with privacy, and data share.

Figure No: 4.1

Source: "Ensuring the Data-Rich Future of the Social Sciences". Gary King, Science Magazine, Vol. 331, 11 February 2011.

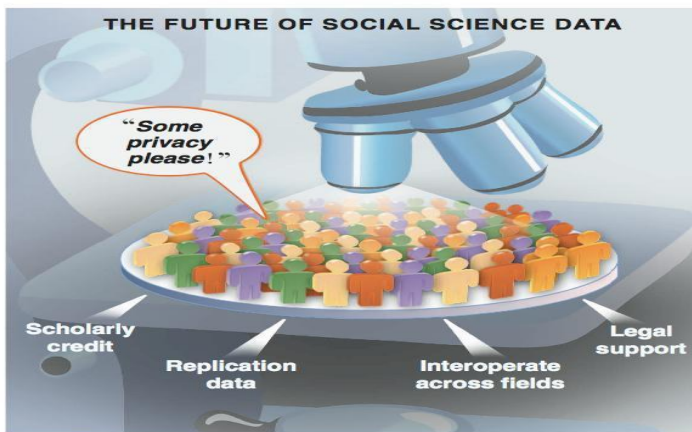


Fig. 1. New types of research data about human behavior and society pose many opportunities if crucial infrastructural challenges are tackled.

### 5. ADDRESSING Vs

#### 5.1 Volume

Unlike most traditional business systems, challenges based data discovery tools are designed to work with an immense number of datasets, so businesses can turn their attention from simply managing the deluge of data to gaining rich insights. From visualizing national marketing campaigns to mining and presenting sales data, the tools enable businesses to derive meaning from large, and growing, volumes of data.

#### 5.2 Variety

Challenges based data discovery tools are designed to mash up or join as many data sources as needed. That means businesses can derive more meaning from structured data, as well as semi structured and unstructured data sources such as social media and sensor data. Using interactive bubble charts, 3-D data landscapes, treemaps, boxplots, heatmaps, word clouds, and many other types of graphics, businesses can view, interpret, and interact with complex data from a multitude of sources.

#### 5.3 Velocity

With challenges based data discovery tools, businesses can replace batch processing with real-time processing of continually updated data streams. The tools also support the democratization of data discovery, so more people can access real-time data sources such as click streams, and analyze and view the data without having to wait for reports.

#### 5.4 Value

When businesses address the three Vs in parallel, they achieve the fourth V: Value. Challenges-based data discovery tools don't just enable users to create attractive info graphics and heatmaps. They create business value by enabling more workers to gain more insights from more data. Instead of waiting weeks or months for static reports, employees can analyze and visualize real-time data on their own. They can also collaborate with co-workers using online, interactive graphics to generate new ideas and identify previously unseen trends. The risks can be greatly reduced by developing safe frameworks within which businesses can seize the immense opportunities presented by challenges-based data discovery tools.

### 6. DERIVING EXPENDITURE

While Apache Hadoop and other technologies are emerging to support back-end concerns such as storage and processing, challenges based data discovery tools focus on the front end of big data on helping businesses explore the data more easily and understand it more fully.

Challenges based data discovery tools allow business users to mash up disparate data sources to create custom analytical views with flexibility and ease of use that simply didn't exist before.

Advanced analytics are integrated in the tools to support creation of interactive, animated graphics on desktops, as well as on powerful mobile devices. End users can view the graphics on the same devices, or on even smaller mobile devices such as tablets or, in limited cases, smartphones.

#### 6.1 Features of Data Discovery Tools

Enable real-time data analysis

Support real-time creation of dynamic, interactive presentations and reports

Allow end users to interact with data, often on mobile devices

Hold data in-memory, where it is accessible to multiple users

Allow users to share and collaborate securely

## 6.2 Extra Skin

Ability to visualize and explore data in-database as well as in memory

Governance dashboard that displays user activity and data lineage

In memory data compression to enable handling of large datasets without driving up hardware costs

Touch optimization for use with touch enabled mobile devices

IEEE Communications Surveys Tutorials 13 (3) (2011) 311–336.

- [14] Salesforce, <http://www.salesforce.com>.
- [15] J. S. Ward and A. Barker, "Undefined By Data: A Survey of Big Data Definitions," <http://arxiv.org/abs/1309.5821v1>.
- [16] SAP Crystal Solutions, <http://www.crystalreports.com/>.
- [17] F. Schmuck, R. Haskin, GPFS: A Shared-Disk File System for Large Computing Clusters, in: Proceedings of the 1st Conference on File and Storage Technologies (FAST'02), Monterey, USA, 2002, pp. 231–244.
- [18] L. Douglas, "3d data management: Controlling data volume, velocity and variety," Gartner, Retrieved 6 (2001).

## ABOUT AUTHORS

**Pavan Kumar Vadrevu** working as assistant professor in department of Information Technology, at Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh.

**Ravi Kumar Suggala** working as assistant professor in department of Information Technology, at Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh.

**G Tej Varma** working as assistant professor in department of Information Technology, at Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh.

## REFERENCES

- [1] Mark Kerzner and Sujee Maniyam, "Hadoop Illuminated," <https://github.com/hadoop-illuminated/hadoop-book>, 2013, Accessed on Sept. 20, 2015.
- [2] IBM What is big data? - Bringing big data to the enterprise. <http://www-01.ibm.com/software/in/data/bigdata/>, Accessed on Sept. 20, 2015.
- [3] Dan Sommer, Rita L. Sallam, James Richardson, "Emerging technology analysis: Visualization-based data discovery tools," June 17, 2011.
- [4] M. A. Beyer and L. Douglas, "The importance of big data: A definition," Stamford, CT: Gartner, 2012.
- [5] SAP HANA One, <http://www.saphana.com/community/solutions/cloud-info> (2013).
- [6] Tom White, "Hadoop: The definitive guide," O'Reilly Media, Inc., 2012.
- [7] S. Ghemawat, H. Gobioff and ST Leung, "The Google file system," in ACM SIGOPS operating systems review, vol. 37, no. 5, ACM, 2003.
- [8] F. Schomm, F. Stahl, G. Vossen, Marketplaces for data: An initial survey, SIGMOD Record 42 (1) (2013) 15–26.
- [9] IBM Big Data & Analytics Hub, <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>, Accessed on Sept. 20, 2015.
- [10] Intel IT Center, "Intel's IT Manager survey on how organizations are using big data," August 2012.
- [11] IBM Institute for Business Value, in collaboration with SAID Business School at the University of Oxford. "Analytics: The real-world use of big data." 2012.
- [12] P. Russom, Big Data Analytics, TDWI best practices report, The Data Warehousing Institute (TDWI) Research (2011).
- [13] S. Sakr, A. Liu, D. Batista, M. Alomari, A survey of large scale data management approaches in cloud environments,