BOOK REVIEW

B. V. Cadambi
Professor, SIBM Pune
Email:vinodcadambi@sibm.edu

TITLE:
BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK

AUTHORS:
VIKTOR MAYER-SCHONBERGER & KENNETH CUKIER
Publisher: John Murray, 338 Euston Road, London NW1 3BH
www.johnmurray.co.uk
Year of Publication: 2013Pages: 242

REVIEW

Information Technology has extensively reengineered businesses and transformed human lives. While numerous benefits have accrued over the years, mankind had to cope with a number of challenges and impediments too. The current emergence of Big Data appears to be no exception.

The book provides a clear understanding of Big Data, their impact and highlights the critical issues for the coming years. Readers of the book would become informed of likely developments. Awareness could be an enabler to smoothen and enhance the process of a cultural transformation

The authors cite some examples to illustrate the importance of timely collection and analysis of data for actionable insights. The example of the fear of HINI epidemic which alerted the Centers for Disease Control and Prevention serves as a powerful illustration. People who contracted the disease delayed consulting doctor sowing to the perception that they suffered from common cold and cough. Hence timely initiatives to prevent the spread of the disease could not be easily taken. At that time,

Google examined the search terms used and found that the frequency of search terms like cold and cough in a location could be used to identify potential locations where H1N1 crisis could occur. This would hasten the process of initiating action. Other examples are predicting the prices of air tickets, concert tickets and hotel rentals.

Gradually the process of using Big Data i.e. data which can be used on a large scale to extract insights so as to initiate changes in markets, organizations or the relationships between citizens and governments commenced. The processing power of computers and the ease with which large volumes of data are available served as enablers.

The Big Data environment has several ramifications as detailed in the book. The first relates to the volume of data used to undertake studies. Since the nineteenth century, people used the statistical theory of sampling during data analysis mainly owing to the time and cost involved in accessing data. Big Data however, offers opportunities to capitalize on the entire mass of data and improve the quality of insights. For instance, through billions of internet searches, it is possible to predict the spread of flu at one or more specific cities. Small samples however, limit the predictions broadly to the state or the country level. This makes it more difficult to initiate quick actions where required.

When sample data are used, efforts are directed to ensure that data are error free. Great care is taken during measurement or data collection. For example, outliers are carefully tracked and eliminated. This would be almost impossible in the realm of Big Data. This however, is not a limitation. For example, at a refinery plant at BP, several wireless sensors are installed throughout the plant. Data on stress on pipes are collected in real time but such data get distorted owing to heat and electrical machinery. However, since a larger mass of data are available quickly, they could be studied carefully. These data revealed that some types of crude oil are more corrosive than others. This quality problem could not have been alleviated easily using smaller datasets. Users of data therefore need to understand the importance of incorporating more data rather than focus on accuracy.

Very often causes for issues are explored although pinpointing causes is rather difficult. With Big Data however, we may search for patterns and obtain statistical correlations between two or more characteristics without trying to pinpoint cause effect relationships. For example, when a larger number of people in a certain locality search for certain terms relating to flu, the chances of a large number of people in that locality catching flu is greater. When limited data are used, such conclusions cannot be easily reached. Some businesses that have capitalized on correlations are

Amazon, Wal-Mart, chemical plants, refineries and the shipping company UPS. Some of these companies could identify critical signs for failure in the near future. This enabled them to initiate timely but not too early intervention. While cost of unnecessary preventive maintenance was not incurred, failure rate was also kept low. Authors therefore advocate a change in mind set that does not seek cause effect relationships.

Since data that traditionally took years to capture are now available quickly, it is important to explore their utility rather than consider them as irrelevant. The authors refer to this as datafication, a process of taking all possible information and transforming such information into a format to make it quantified. For example, Matthew Fontaine Maury, a US Naval Officer capitalized on all data available from naval logs such as weather, water and wind and maps, charts and nautical books. He also collected further data such as speed, direction of winds and waves, the months of travel. Patterns were identified and charts were developed to guide voyages. This helped to reduce the travel time drastically. Other examples of datafication are include (i) development of anti-theft systems in cars using data on a seated person's posture, distribution of weight and other details so that only authorized individuals drive the car (ii) Global positioning systems (GPS) to receive signals from overhead satellites (iii) Social networking platforms.

Historically, a need for data and its use as a support in managing business has been recognized. However, lately the scope to capitalize on data has increased considerably and evidences of successful use are discernible. For example, Amazon records not only the books that customers buy, but also other books that customers peruse. This information has helped Amazon to suggest suitable titles to customers. As a result, customer experience is enhanced. Similarly, IBM has developed predictive models for users of battery operated cars. Models incorporate data on the battery, location of cars, timing of the day and historical patterns of usage, etc. Models guide users in identifying optimal points in time to recharge batteries incorporating relevant information.

Thus, there is scope to use data in more ambitious ways than envisaged at first sight. To capitalize on data, it would be better to draw data from multiple sources. Further, the scope to provide for repetitive use of data should be provided for. In several countries, like USA, UK, Australia and Brazil, governments have therefore made data available for public use.

Besides drawing on data, it is important to unleash its power using innovative approaches. To effectively enable this, three types of Big Data companies have evolved and each type has varied expertise. The first type refers to companies that

possess data or those that have ready access to it. The next relates to companies that draw on such data to extract insights. Finally, the third type of companies work on ideas for others to use such data. It is important to recognize the power of the entire value chain. For example, airlines provide information on tickets sold to Farecast who analyzed the buying patterns to help customers to plan the timing of purchase. Farecast shared the commission with airlines, leading to a win-win situation. Groups that work on ideas to use data are free from biases that could otherwise deter spotting opportunities.

While Big Data offers extensive promise, important challenges need to be tactfully addressed. Extensive data on individuals are captured from public sources and used by others which could result in a threat to privacy. For example, Google may capture extensive information on individuals who expect to contract flu. It is not easy to ensure anonymity. Further, such data are used at varied points in time and it is difficult to get the consent of people. Certain data should therefore be used with adequate safeguards. Regulators should establish ground rules for users to assess potential dangers and identify ways to mitigate their impact. Thus, the responsibility to protect privacy has to be transferred to data users.

Big Data are sometimes used to predict potential suspects who are likely to create major problems. It may be tempting to take action on such individuals in advance to preempt catastrophic situations. Instead, people could be held for past actions rather than anticipated actions.

When complex patterns are studied, algorithms are used without an in-depth understanding of the logic. This however, should not be unduly unfavorable to individuals such as potential suspects. The authors assert that it is expected that a new breed of professionals called "algorithmists" will emerge. These professionals will be experts in the areas of computer science, mathematics and statistics and they would review the analysis and predictions from big data. Algorthimists need to be both impartial and confidential.

Despite the challenges, Big Data interventions are critical. They are particularly so in addressing critical concerns for mankind such as healthcare, crimes and calamities.

The authors have done a commendable job in articulating various issues relating to the subject. Apt chapter titles have been assigned. For example, a chapter describing the scope to use a large mass of data is titled "More". Similarly the importance of using all data without excessive concern for correctness is termed "Messy". Examples from diverse business settings are included. The author draws on parallels from History which serve as useful lessons and offer insights for the future.

Readers would certainly benefit as they would be better prepared to cope with future demands. Organizations and governments could in particular, address important demands on reengineering and initiate steps like manpower planning as the demand for data analysts and other professionals will grow.