

The Evolution of Big Data

Hiren H Darji¹

¹Assistant Professor

¹ Anand Institute of Information Science, Anand, India

Abstract— this paper provides brief idea about Big Data, various sources which generate rich amount of Big Data and how Big Data are analyzed by using various tools or technology. Also this paper briefly describes three very important characteristics about Big Data. This paper also includes survey of Big Data in Global and Financial/Banking Sector.

Key words: Big Data, Volume, Velocity, Variety, Global, Finance.

I. INTRODUCTION

Advances in digital sensors, communications, computation, and storage have created huge collections of data, capturing information of value to business, science, government, and society. For example, search engine companies such as Google, Yahoo!, and Microsoft have created an entirely new business by capturing the information freely available on the World Wide Web and providing it to people in useful ways. These companies collect trillions of bytes of data every day and continually add new services such as satellite images, driving directions, and image retrieval. The societal benefits of these services are immeasurable, having transformed how people find and make use of information on a daily basis.

Just as search engines have transformed how we access information, other forms of big data computing can and will transform the activities of companies, scientific researchers, medical practitioners, and our nation's defense and intelligence operations. Some examples include:

- 1) Modern medicine collects huge amounts of information about patients through imaging technology (CAT scans, MRI), genetic analysis (DNA microarrays), and other forms of diagnostic equipment. By applying data mining to data sets for large numbers of patients, medical researchers are gaining fundamental insights into the genetic and environmental causes of diseases, and creating more effective means of diagnosis.
- 2) The collection of all documents on the World Wide Web (several hundred trillion bytes of text) is proving to be a corpus that can be mined and processed in many different ways. [5]

II. BIG-DATA TECHNOLOGY: SENSE, COLLECT, STORE, AND ANALYZE

A. Sensors

Digital data are being generated by many different sources, including digital imagers (telescopes, video cameras, MRI machines), chemical and biological sensors (microarrays, environmental monitors), and even the millions of individuals and organizations generating web pages.

B. Computer networks

Data from the many different sources can be collected into massive data sets via localized sensor networks, as well as the Internet.

C. Data storage

Advances in magnetic disk technology have dramatically decreased the cost of storing data.

D. Cluster computer systems

A new form of computer systems, consisting of thousands of "Nodes," each having several processors and disks, connected by high-speed local-area Networks, These clusters provide both the storage capacity for large data sets, and the computing power to organize the data, to analyze it, and to respond to queries about the data from remote users.

E. Cloud computing facilities

The rise of large data centers and cluster computers has created a new business model, where businesses and individuals can rent storage and computing capacity, rather than making the large capital investments needed to construct and provision large-scale computer installations. [5]

III. BIG DATA INDUSTRIAL REVOLUTION

The enormous volumes of data require automated or semi-automated analysis – techniques to detect patterns, identify anomalies, and extract knowledge. [3]

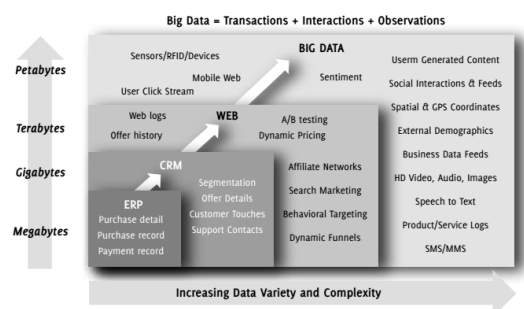


Fig. 1: Industrial Revolution

IV. THE IMPORTANCE OF BIG DATA

The reason why we should wish to have and examine all that data is evident. The social media, web analytics, log files, sensors, and suchlike all provide valuable information, while the cost of it solutions continues to drop and computer-processing power is increasing. With developments like these, the surplus of information seems to have largely vanished: in principle, organizations are now capable of managing the flood of data and to use it to their own (financial) advantage. Those who excel in acquiring,

processing, and managing valuable data, says Gartner, will be able to realize a 20% better result, in financial terms, than their competitors.

Within organizations, the share of unstructured data, such as documents, e-mail and images, is around 60 to 80 per cent. Of all data analyses that currently take place in organizations, 5 to 15 per cent contain a social component that enriches the structured data. This number must increase, not least because of all the external data that can be included in the analyses. [3]

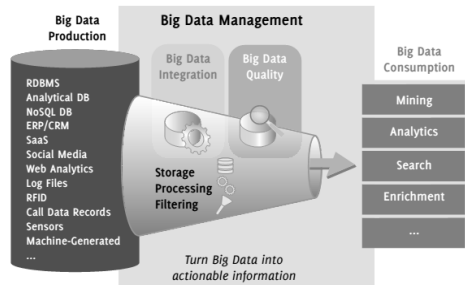


Fig. 2: Big Data Management

V. VOLUME, VARIETY, VELOCITY

Three factors can influence one another in the growth of data flow: the quantity of data (Volume), the nature of the data type: structured, semi-structured and unstructured (Variety) and the desired analysis rate (Velocity). Nowadays we often add Complexity, Value and Relevance to this list. The last two are included because we would like to know what we can and want to do with all the data, so that we are not investing time, money and effort for no return. [3]

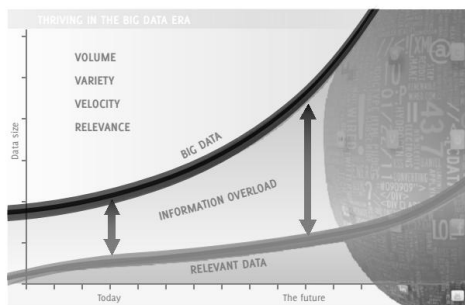


Fig. 3: ratio of Volume, Variety, Velocity

VI. BIG DATA IN ORGANIZATIONS IN THE YEAR 2012

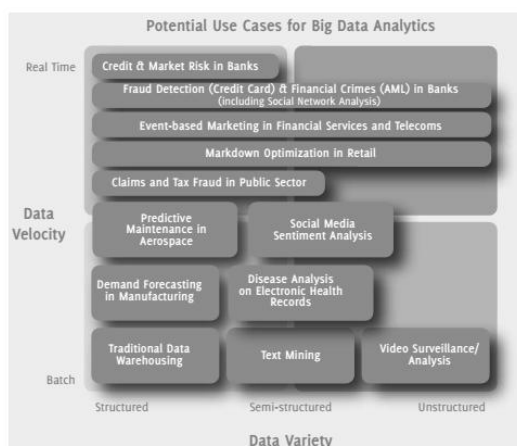


Fig. 4: Data Variety

A. Tools used in Big Data Scenario

1) NoSQL

Databases MongoDB, CouchDB, Cassandra, Redis, BigTable, Hbase, Hypertable, Voldemort, Risk, ZooKeeper

2) Storage

S3, Hadoop Distributed File System

3) Servers

EC2, Google App Engine, Elastic, Beanstalk, Heroku

4) Processing

R, Yahoo! Pipes, Mechanical Turk, Solr/Lucene, ElasticSearch, Data Meer, Big Sheets, Tinker pop

B. Big data adoption for Global Market and Banking/Finance Sector



Source: Analytics: The real-world use of big data, a collaborative research study by the IBM Institute for Business Value and the Saïd Business School at the University of Oxford. © IBM 2012

Fig. 5: Storage

1) Educate

Building a base of knowledge: 26 percent of banking and financial markets respondents

2) Explore

Defining the business case and roadmap: 47 percent of banking and financial markets respondents

3) Engage

Embracing big data: 23 percent of banking and financial markets respondents

4) Execute

Implementing big data at scale: 3 percent of banking and financial markets respondents. [2]

VII. CONCLUSION

By reading this paper a reader gets fair idea about the concept Big Data, various tools which are used in Big Data analysis and by using information provided in this paper, reader gets interested in new research area called Big Data.

REFERENCES

- [1] A community white paper "Challenges and Opportunities with Big Data".
- [2] Analytics: The real-world use of big data in financial services.
- [3] Jaap Bloem, Menno van Doorn, Thomas van Manen "Creating clarity with Big Data".
- [4] Marko Grobelnik, Jozef Stefan Institute "Big Data Tutorial".
- [5] Randy H. Katz, University of California, and Berkeley "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society".
- [6] SAS, "2013 Big Data Survey Research"
- [7] World Economic Forum "Big Data, Big Impact: New Possibilities for International Development"