Volume 6, Issue 7, July 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Big Data: The New Era of Storing Data

Salil Jagtap
Computer Science and Engineering,
SGBAU, Maharashtra, India

Shraddha Malviya
Electronics and Telecommunication Engg,
SGBAU, Maharashtra, India

Abstract*— *we have entered the big data era. Organizations are capturing, storing, and analysing data that has high volume, velocity, and variety and comes from a variety of new sources, including social media, machines, log files, video, text, image, RFID, and GPS. These sources have strained the capabilities of traditional relational database management systems and spawned a host of new technologies, approaches, and platforms. The potential value of big data analytics is great and is clearly established by a growing number of studies. The keys to success with big data analytics include a clear business need, strong committed sponsorship, alignment between the business and IT strategies, a fact-based decision-making culture, a strong data infrastructure, the right analytical tools, and people skilled in the use of analytics. Because of the paradigm shift in the kinds of data being analysed and how this data is used, big data can be considered to be a new, fourth generation of decision support data management.

Keywords*— *Big Data, Business intelligence, Analytics, Hadoop, MapReduce.

I. INTRODUCTION

Data becomes big data when its volume, velocity or variety exceeds the abilities of your IT systems to ingest, store, analyse and process it. Many organizations have the equipment and expertise to handle large quantities of structured data—but with the increasing volume and faster flows of data, they lack the ability to mine it and derive actionable intelligence in a timely way. Not only is the volume of this data growing too fast for traditional analytics, but the speed with which it arrives and the variety of data types necessitates new types of data processing and analytic solutions.
However, big data doesn't always fit into neat tables of columns and rows. There are many new data types, both structured and unstructured, that can be processed to yield insight into a business or condition. There is a need to reduce the data to the point that it can be put in a structured form. Then it can be meaningfully compared to the rest of your data, and scrutinized with traditional business intelligence (BI) tools.

II. BIG DATA CONCEPT

Big Data is defined by its size, comprising a large, complex and independent collection of data sets, each with the potential to interact. In addition, an important aspect of Big Data is the fact that it cannot be handled with standard data management techniques due to the inconsistency and unpredictability of the possible combinations.
In IBM's view Big Data has four aspects:

A. VOLUME.

It implies the amount of data. While volume indicates more data, it is the granular nature of the data that is unique. Big data requires processing high volumes of low-density, unstructured Hadoop data—that is, data of unknown value, such as Twitter data feeds, click streams on a web page and a mobile app, network traffic, sensor-enabled equipment capturing data at the speed of light, and many more. It is the task of big data to convert such Hadoop data into valuable information. For some organizations, this might be tens of terabytes, for others it may be hundreds of petabytes.

B. VELOCITY

It implies the rate at which data is received and perhaps acted upon. The highest velocity data normally streams directly into memory versus being written to disk. Some Internet of Things (IoT) applications have health and safety ramifications that require real-time evaluation and action. Other internet-enabled smart products operate in real time or near real time. For example, consumer e-Commerce applications seek to combine mobile device location and personal preferences to make time-sensitive marketing offers. Operationally mobile application experiences have large user populations, increased network traffic and the expectation for immediate response.

C. VARIETY

It implies new unstructured data types. Unstructured and semi-structured data types such as text, audio, and video require additional processing to both derive meaning and the supporting metadata. Once understood, unstructured data has many of the same requirements as structured data, such as summarization, lineage, auditability, and privacy. Further complexity arises when data from a known source changes without notice. Frequent or real-time changes are an enormous burden for both transaction and analytical environments.

D.VALUE

Data has intrinsic value but it must be discovered. There are a range of quantitative and investigative techniques to derive value from data from discovering a consumer preference or sentiment, to making a relevant offer by location or for identifying a piece of equipment that is about to fail. The technological breakthrough is that the cost of data storage and compute has exponentially decreased, thus providing an abundance of data from which statistical analysis on the entire data set versus previously only sample. The technological breakthrough makes much more accurate and precise decisions possible. However, finding value also requires new discovery processes involving clever and insightful analysts, business users, and executives. The real big data challenge is a human one, which is learning to ask the right questions, recognizing patterns, making informed assumptions, and predicting behaviour.

III. IMPACT OF BIG DATA ON BUSINESS ENTERPRISE

Data are generated in a growing number of ways. Use of traditional transactional databases has been supplemented by multimedia content, social media and myriad types of sensors. Advances in information technology allow users to capture, communicate, aggregate, store and analyse enormous pools of data known as BIG DATA.

However, the new data collection methodologies pose a dilemma for businesses that have depended upon database Technology to store and process data. Big data derives its name from the fact that the datasets are large enough that typical database systems are unable to capture, save, and analyse these datasets. The actual size of big data varies by business sector, software tools available in the sector, and average dataset sizes within the sector. Best estimates of size range from a few dozen terabytes to many petabytes. In order to benefit from big data, new storage technologies and analysis methods need to be adopted. Business executives must determine the new technologies and methodologies best suited to their information needs. Business executives ignoring the growing field of big data will eventually become non-competitive.

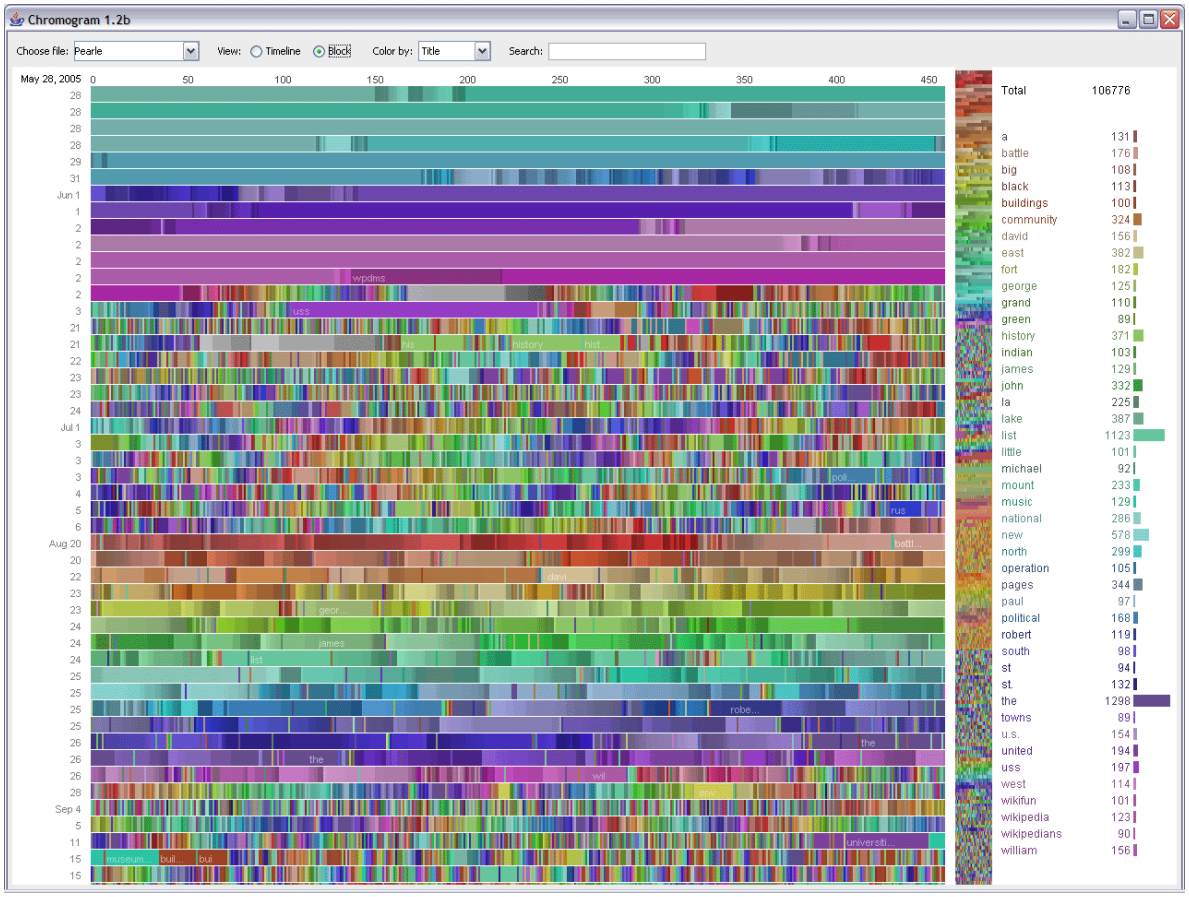

Fig1. Multiple terabyte sized example of Big Data

IV. BIG DATA ANALYTICS

Big data analytics is the process of collecting, organizing and analysing large sets of data (called big data) to discover patterns and other useful information. Big data analytics can help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and future business decisions. Analysts working with big data basically want the knowledge that comes from analysing the data. To analyse such a large volume of data, big data analytics is typically performed using specialized software tools and applications for predictive analytics, data mining, text mining, and forecasting and data optimization. Collectively these processes are separate but highly integrated functions of high-performance analytics. Using big data tools and software enables an organization to process extremely large volumes of data that a business has collected to determine which data is relevant and can be analysed to drive better business decisions in the future.

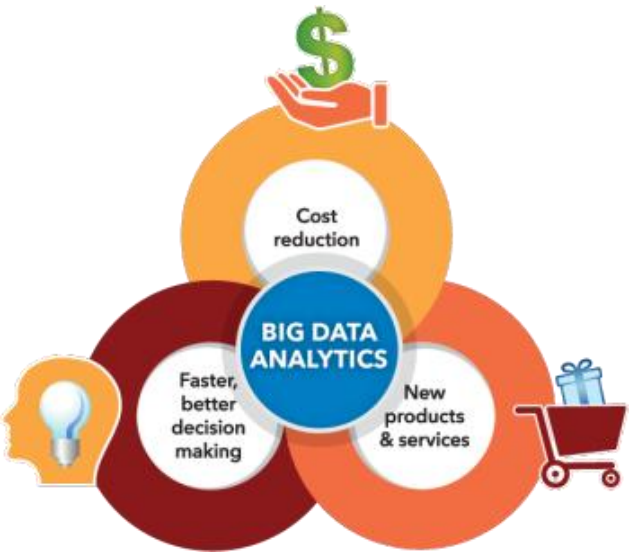

Fig 2. Big data analytics

V. BIG DATA TECHNOLOGIES

Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, cost reductions, and reduced risks for the business. To harness the power of big data, we would require an infrastructure that can manage and process huge volumes of structured and unstructured data in real-time and can protect data privacy and security. There are various technologies in the market from different vendors including Amazon, IBM, Microsoft, etc to handle big data. While looking into the technologies that handle big data, we examine the following two classes of technology:

A. OPERATIONAL BIG DATA

For operational Big Data workloads, NoSQL Big Data systems such as document databases have emerged to address a broad set of applications and other architectures such as key-value stores, column family stores, and graph databases are optimized for more specific applications. NoSQL technologies, which were developed to address the shortcomings of relational databases in the modern computing environment, are faster and scale much more quickly and inexpensively than relational databases.

Critically, NoSQL Big Data systems are designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently. This makes operational Big Data workloads much easier to manage and cheaper and faster to implement.

In addition to user interactions with data, most operational systems need to provide some degree of real-time intelligence about the active data in the system. For example in a multi-user game or financial application, aggregates for user activities or instrument performance are displayed to users to inform their next actions. Some NoSQL systems can provide insights into patterns and trends based on real-time data with minimal coding and without the need for data scientists and additional infrastructure.

B. ANALYTICAL BIG DATA

Analytical Big Data workloads on the other hand tend to be addressed by MPP database systems and MapReduce. These technologies are also a reaction to the limitations of traditional relational databases and their lack of ability to scale beyond the resources of a single server. Furthermore, MapReduce provides a new method of analysing data that is complementary to the capabilities provided by SQL.

As applications gain traction and their users generate increasing volumes of data, there are a number of retrospective analytical workloads that provide real value to the business. Where these workloads involve algorithms that are more sophisticated than simple aggregation, MapReduce has emerged as the first choice for Big Data analytics. Some NoSQL systems provide native MapReduce functionality that allows for analytics to be performed on operational data in place. Alternately, data can be copied from NoSQL systems into analytical systems such as Hadoop for MapReduce.

VI. BENEFITS AND IMPORTANCE OF BIG DATA

The main importance of Big Data consists in the potential to improve efficiency in the context of use a large volume of data, of different type. If Big Data is defined properly and used accordingly, organizations can get a better view on their business therefore leading to efficiency in different areas like sales, improving the manufactured product and so forth. Using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums. Using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production. Using the data regarding the previous medical history of patients, hospitals are providing better and quick service.

VII. CONCLUSION

Organizations in every industry are trying to make sense of the massive influx of big data, as well as to develop analytic platforms that can synthesize traditional structured data with semi-structured and unstructured sources of information. When properly captured and analysed, big data can provide unique insights into market trends, equipment failures, buying patterns, maintenance cycles and many other business issues, lowering costs, and enabling more targeted business decisions. To obtain value from big data, you need a cohesive set of solutions for capturing, processing, and analysing the data, from acquiring the data and discovering new insights to making repeatable decisions and scaling the associated information systems.

REFERENCES

[1] The Intel science and technology centre for big data, http://istc-bigdata.org.
[2] bigdata@csail, http://bigdata.csail.mit.edu/.
[3] The Intel science and technology centre for big data, http://istc-bigdata.org.
[4] Apache Hadoop, http://hadoop.apache.org.
[5] IBM SmartCloud Enterprise,http://www-935.ibm.com/services/us/en/cloud-enterprise/ (2012).
[6] D. Lazer, R. Kennedy, G. King, A. Vespignani, The Parable of Google Flu: Traps in Big Data Analysis, Science 343 (2014) 1203–1205.
[7] D. J. Abadi, Data Management in the Cloud: Limitations and Opportunities, IEEE Data Engineering Bulletin 32 (1) (2009) 3–12.
[8] https://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf
[9] https://www.ida.gov.sg/~/media/Files/Infocomm%20Landscape/Technology/TechnologyRoadmap/BigData pdf
[10] https://www.researchtrends.com/wp-content/uploads/2012/09/Research_Trends_Issue30.pdf

