International Journal of Computer Applications Technology and Research
Volume 5–Issue 7, 460-464, 2016, ISSN:2319–8656

Big Data Analytics: Recent Achievements and New Challenges

Dr Anand Mohan
NSHM Group of Institutions,
Durgapur-713212(West Bengal) India

Abstract : The era of Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills. Big data has become an important issue for a large number of research areas such as data mining, machine learning, computational intelligence, information fusion, the semantic Web, and social networks. The combination of big data technologies and traditional machine learning algorithms has generated new and interesting challenges in other areas as social media and social networks. These new challenges are focused mainly on problems such as data processing, data storage, data representation, and how data can be used for pattern mining, analysing user behaviours, and visualizing and tracking data, among others. In this paper, discussion about the new concept big data and data analytic their concept, tools and methodologies that is designed to allow for efficient data mining and information sharing fusion from social media and of the new applications and frameworks that are currently appearing under the "umbrella" of the social networks, social media and big data paradigms.

Keywords - Big data, Data mining, Social media, Social networks, Social-based frameworks and applications

INTRODUCTION

Big Data is a term used to describe a massive volume of diverse data, both structured and unstructured, that is so large and fast-moving that it's difficult or impossible to process using traditional databases and software technology. In most enterprise scenarios, the data is too enormous, streaming by too quickly at unpredictable and variable speeds, and exceeds current processing capacity. According to leading technology research firm Gartner Inc., Big Data is high-volume, high-velocity, and high-variety of information assets that demand cost-effective, innovative forms of information processing for enhanced insights and decision-making. While Big Data is defined by its characteristics, the 3 "Vs" (i.e., volume, velocity, and variety), other

analysts add a fourth "V" to represent the data's value. Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges that we face with DBMS tools and other technologies is capture, storage, search, sharing, transfer, analysis, and visualization.

Key enablers for the appearance and growth of 'Big-Data' are:

- Increase in storage capabilities
- Increase in processing power
- Availability of data

Globally Big Data generation in Facebook creates over 30 billion pieces of content per day and stores 30 petabytes of data and Twitter produces over 90 million tweets per day[1].

The term **"Big Data Analytics"**, when used by software vendors, refers to the technology that an organization requires to handle data at extreme scales. Not only does this make Big Data management and storage vastly different from normal or structured data that most people are accustomed to handling, but it also means that organizations now require powerful, integrated solutions for making this information usable and applicable for business analytics practices and dealing with large data sets, organizations face difficulties in being able to integrate, manipulate, and manage them efficiently and effectively.

Big Data is a huge problem in business analytics because standard tools and procedures are not designed to search and analyze massive data sets. The use of Big Data large pools of data that can be brought together and analyzed to discern patterns and make better decisions will become the basis of competition and growth for individual firms, enhancing productivity and creating significant value for the world economy by reducing waste and increasing the quality of products and services[2].

Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information. Big

data can be characterized by 3Vs (volume, variety and velocity) are three defining properties or dimensions of big data. Volume refers to the amount of data, variety refers to the number of types of data and velocity refers to the speed of data processing.

According to the 3Vs (volume, variety and velocity) model, the challenges of big data management result from the expansion of all three properties, rather than just the volume alone - the sheer amount of data to be managed. The extreme volume of data, the wide variety of types of data and the velocity at which the data must be must processed. Although big data doesn't refer to any specific quantity, the term is often used when speaking about petabytes (A petabyte is a measure of memory or storage capacity and is 2 to the 50th power bytes or, in decimal, approximately a thousand terabytes) and exabytes (An exabyte (EB) is a large unit of computer data storage, two to the sixtieth power bytes. The prefix *exa* means one billion, or one quintillion, which is a decimal term. Two to the sixtieth power is actually 1,152,921,504, 606,846,976 bytes in decimal, or somewhat over a quintillion (or ten to the eighteenth power) bytes. It is common to say that an exabyte is approximately one quintillion bytes. In decimal terms, an exabyte is a billion gigabytes of data, much of which cannot be integrated easily).

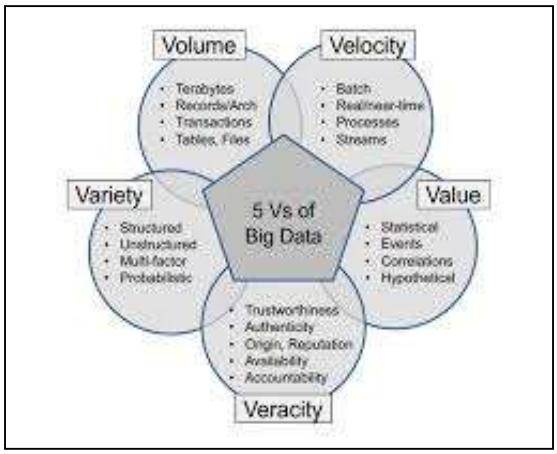

Fig 1-5Vs of Big Data

Big Data will help to create new growth opportunities and entirely new categories of companies, such as those that aggregate and analyse industry data. Many of these will be companies that sit in the middle of large information flows where data about products and services, buyers and suppliers, consumer preferences and intent can be captured and analysed. Forward-thinking leaders across sectors should begin aggressively to build their organisations' Big Data capabilities[3].

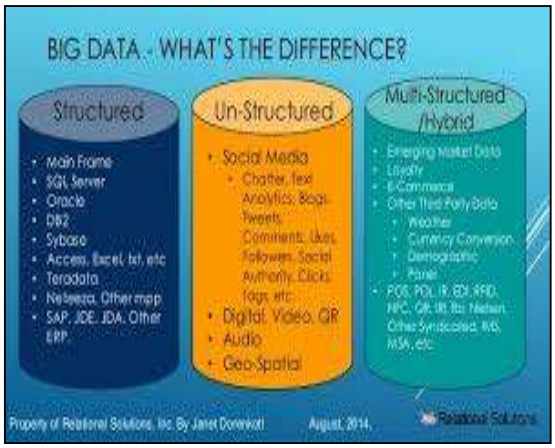

Fig 2 - Big Data Structured, Un-Structured and Multi Structured/Hybrid

Five ways to leverage Big Data are given as :

1. Big Data can unlock significant value by making information transparent. There is still a significant amount of information that is not yet captured in digital form, e.g., data that are on paper, or not made easily accessible and searchable through networks. We found that up to 25 percent of the effort in some knowledge worker workgroups consists of searching for data and then transferring them to another (sometimes virtual) location. This effort represents a significant source of inefficiency.

2. As organisations create and store more transactional data in digital form, they can collect more accurate and detailed performance information on everything from product inventories to sick days and therefore expose variability and boost performance. In fact, some leading companies are using their ability to collect and analyse big data to conduct controlled experiments to make better management decisions.

3. Big Data allows ever-narrower segmentation of customers and therefore much more precisely tailored products or services.

4. Sophisticated analytics can substantially improve decision-making, minimise risks, and unearth valuable insights that would otherwise remain hidden.

5. Big Data can be used to develop the next generation of products and services. For instance, manufacturers are using data obtained from sensors embedded in products to create innovative after-sales service offerings such as proactive maintenance to avoid failures in new products.

Big data expands the possible domains of application for algorithms and machine-mediated analysis. At some manufacturers, for example, algorithms analyze sensor data from production lines, creating self-regulating processes that cut waste, avoid costly (and sometimes dangerous) human interventions, and ultimately lift output. In advanced, "digital" oil fields, instruments constantly read data on wellhead conditions, pipelines, and mechanical systems. That information is analyzed by clusters of computers, which feed their results to real-time operations centers that adjust oil flows to optimize production and minimize downtimes. One major oil company has cut operating and staffing costs by 10 to 25 percent, while increasing production by 5 percent[4].

Computer scientists, physicists, economists, mathematicians, political scientists, bio-informatics, sociologists, and other scholars are clamouring for access to the massive quantities of information produced by and about people, things, and their interactions.

Benefits of big data analytics tools are given below :

1. Big data and big data tools offer many benefits. The main business advantages of big data generally fall into one of three categories: cost savings, competitive advantage, or new business opportunities.
2. **Cost Savings**
3. Big data tools like Hadoop allow businesses to store massive volumes of data at a much cheaper price tag than a traditional database. Companies utilizing big data tools for this benefit typically use Hadoop clusters to augment their current data warehouse, storing long-term data in Hadoop rather than expanding the data warehouse. Data is then moved from Hadoop to the traditional database for production and analysis as needed. Versatile big data tools can also function as multiple tools at once, saving organizations on the cost of needing to purchase more tools for the same tasks.
4. **Competitive Advantage**
5. New Business Opportunities
6. The final benefit of big data analytics tools is the possibility of exploring new business opportunities. Entrepreneurs have taken advantage of big data technology to offer new services in AdTech and MarketingTech. Mature companies can also take advantage of the data they collect to offer add-on services or to create new product segments that offer additional value to their current customers. In addition to those benefits, big data analytics can pinpoint new or potential audiences that have yet to be tapped by the enterprise. Finding whole new customer segments can lead to tremendous new value.
7. These are just a few of the actionable insights made possible by available big data analytics tools. Whether an organization is looking to boost sales and marketing results, uncover new revenue opportunities, improve customer service, optimize operational efficiency, reduce risk, improve security, or drive other business results, big data insights can help[5].

The use cases for big data analysis are given below :

1. Big data analytics lends itself well to a large variety of use cases spread across multiple industries. Financial institutions can quickly find that big data analysis is adept at identifying fraud before it becomes widespread, preventing further damage. Governments have turned to big data analytics to increase their security and combat outside cyber threats. The healthcare industry uses big data to improve patient care and discover better ways to manage resources and personnel. Telecommunications companies and others utilize big data analytics to prevent customer churn while also planning the best ways to optimize new and existing wireless networks. Marketers have quite a few ways they can use big data. One involves sentiment analysis, where marketers can collect data on how customers feel about certain products and services by analyzing what consumers post on social media sites like Facebook and Twitter.
2. The number of use cases are plentiful, and no industry should think that analytics couldn't be used in some way to improve their businesses. That type of versatility is part of what has made big data so popular. And these are only a few examples of use cases. As companies and other organizations become more familiar with all of the capabilities granted through big data analytics, more use cases will likely be discovered, adding to big data's overall value. As with any developing technology, the process may take some time, but eventually its widespread use will lead to the discovery of even more benefits and uses.

Some of the Top Big Data Tools overview:

1. **Apache Hadoop :**Hadoop is an open source software framework originally developed by Doug Cutting and Mike Cafarella in 2006. It was specifically built to handle very large data sets. Hadoop is made up of two main parts: the Hadoop Distributed File System (HDFS) and MapReduce. HDFS is the storage component of Hadoop. Hadoop stores data by splitting files into large blocks and distributing it across nodes. MapReduce is the processing engine of Hadoop.

Hadoop processes data by delivering code to nodes to process in parallel.

2. **Apache Spark :**Apache Spark is quickly growing as a data analytics tool. It is an open source framework for cluster computing. Spark is frequently used as an alternate to Hadoop's MapReduce because it is able to analyze data up to 100 times faster for certain applications. Common use cases for Apache Spark include streaming data, machine learning and interactive analysis.

3. **Apache Hive :**Apache Hive is a SQL-on-Hadoop data processing engine. Apache Hive excels at batch processing of ETL jobs and SQL queries. Hive utilizes a query language called HiveQL. HiveQL is based on SQL, but does not strictly follow the SQL-92 standard.

4. **NoSQL Databases :** NoSQL databases have grown in popularity. These Not Only SQL databases are not bound by traditional schema models allowing them to collect unstructured datasets. The flexibility of NoSQL databases like MongoDB, Cassandra, and HBase make them a popular option for big data analytics[6].

Big data in the cloud - Big data analytics can be a complex concept, one that many businesses may feel like they're not ready for. Big data infrastructure can get to be complicated, and without the right personnel on hand, maintaining it can be a monumental task. One solution to this significant problem is for companies to head to the cloud for their big data needs. Many cloud vendors already provide a variety of services through the cloud, and big data analytics is just the latest example of this. Taking big data to the cloud offers up a number of advantages. Improvements come in the form of better performance, targeted cloud optimizations, more reliability, and greater value. Big data in the cloud gives businesses the type of organizational scale many are searching for. This allows many users, sometimes in the hundreds, to query data while only being overseen by a single administrator. That means little supervision is required.

Big data in the cloud also allows organizations to scale quickly and easily. This scaling is done according to the customer's workload. If more clusters are needed, the cloud can give them the extra boost. During times of less activity, everything can be scaled down. This added flexibility is particularly valuable for companies that experience varying peak times. Big data in the cloud also takes advantage of the benefits of cloud infrastructure, whether they be from Amazon Web Services, Microsoft Azure, Google Cloud Platform, or others.

Gathering data from various sources is, of course, only one part of the big data analytics process. All that data needs to be stored somewhere, and that repository is often referred to as a **data lake**. Data lakes are where data is kept in its raw form, before any organizational structure is used and before any analytics are performed. Data lakes don't use the traditional structure of files or folders but rather use a flat architecture where each element has its own identifier, making it easy to find when queried. Data lakes are a type of object storage that Hadoop uses, making it an effective way to describe where Hadoop-supported platforms pull their data from. One major benefit of having a data lake is the ability to store massive amounts of data. As big data continues to grow, the need for that near limitless storage capability has grown with it. Data lakes also allow for added processing power while also providing the ability to handle numerous jobs at the same time. These are all capabilities that have been increasingly in demand as more enterprises use big data analytics tools.

Many different types of solutions are required to support the wide range of big data use cases. From simple spreadsheets to advanced analytics and marketing solutions to analytics engines, Qubole provides effortless integration to centrally analyze your data all in one spot.

Spreadsheets and Analytics Tools: Through ODBC connectors, Qubole customers can connect to Microsoft Excel and tools from leading analytics vendors such as Tableau, Qlik, MicroStrategy, and TIBCO Jaspersoft. In addition, the R statistical programming language can be integrated with Qubole using ODBC/REST APIs.

Analytics Engines: Qubole offers connectors for massively parallel processing databases such as Vertica as well as relational database engines such as Microsoft SQL Server and the MySQL open source database, and NoSQL databases such as MongoDB[7].

CRM and Online Marketing Solutions: Qubole also connects to leading CRM and online marketing platforms such as Salesforce.com and online marketing and web analytics solutions such as Omniture and Google Analytics.

CONCLUSION

The era of Big Data could yield new management principles. In the early days of professionalized corporate

management, leaders discovered that minimum efficient scale was a key determinant of competitive success. Likewise, future competitive benefits are likely to accrue to companies that can not only capture more and better data but also use that data effectively at scale. Companies have decided that big data is not just a buzzword, but a new fact of business life -- one that requires having strategies in place for managing large volumes of both structured and unstructured data and with the reality of big data comes the challenge of analyzing it in a way that brings real business value. Business and IT leaders who started by addressing big data management issues are now looking to use big data analytics to identify trends, detect patterns and glean other valuable findings from the sea of information available to them. Big data analytics technologies on their own aren't sufficient to handle the task. Well-planned analytical processes and people with the talent and skills needed to leverage the technologies are essential to carry out an effective big data analytics initiative. The information resources collected here to learn about big data analytics best practices from experienced users and industry analysts -- from identifying business goals to selecting the best big data analytics tools for organization's needs.

REFERENCES

[1]. http://searchcloudcomputing.techtarget. com
[2].https://www.qubole.com/resources/articles/big-data-analytics/?nabe= 5695374637924352:1&utm_referrer=https%3A%2F%2Fwww.google.co.in#sthash.UL6ZrfK9.dpuf
[3].http://searchbusinessanalytics.techtarget.com/essentialguide/Guide-to-big-data-analytics-tools-trends-and-best-practices
[4].http://www.infoworld.com/article/2616959/big-data/7-top-tools-for-taming-big-data.html
[5].http://iveybusinessjournal.com/publication/why-big-data-is-the-new-competitive-advantage/
[6].http://www.in.techradar.com/news/world-of-tech/The-importance-of-big-data-analytics-in-business/articleshow/44104837.cms
[7].http://datascienceseries.com/stories/ten-practical-big-data-benefits