BIG DATA ANALYTICS

Mr. A. KAMALRAJ
Assistant Professor in Information Technology,
AJK College of Arts and Science,
Navakkarai, Coimbatore – 641 105.

Abstract

Big data is a new driver of the world economic and societal changes. The world's data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Big Data Analytics technologies. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Potential breakthroughs include new algorithms, methodologies, systems and applications in Big Data Analytics that discover useful and hidden knowledge from the Big Data efficiently and effectively. Big data analytics must also be team effort cutting across academic institutions, government and society and industry and by researchers from multiple disciplines including computer science and engineering, health, data science and social and policy areas.

Keywords: big data, analytics, evaluation, industries etc.,

INTRODUCTION

The concept of big data has been around for years; most organizations now understand that if they capture all the data that streams into their businesses, they can apply analytics and get significant value from it. But even in the 1950s, decades before anyone uttered the term "big data," businesses were using basic analytics(essentially numbers in a spreadsheet that were manually examined) to uncover insights and trends.

The new benefits that big data analytics brings to the table, however, are speed and efficiency. Whereas a few years ago a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before.

Big Data analytics is the process of collecting, organizing and analyzing 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 typically want the *knowledge* that comes from analyzing the data.

WHY IS BIG DATA ANALYTICS IMPORTANT?

Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, higher profits and happier customers.

1. **Cost reduction.** Big data technologies such as Hadoop and cloud-based analytics bring significant cost advantages when it comes to storing large amounts of data – plus they can identify more efficient ways of doing business.
2. **Faster, better decision making.** With the speed of Hadoop and in-memory analytics, combined with the ability to analyze new sources of data, businesses are able to analyze information immediately – and make decisions based on what they've learned.
3. **New products and services.** With the ability to gauge customer needs and satisfaction through analytics comes the power to give customers what they want. Davenport points out that with big data analytics, more companies are creating new products to meet customers' needs.

WHO'S USING IT?

Travels and Hospitality

Keeping customers happy is key to the travel and hotel industry, but customer satisfaction can be hard to gauge – especially in a timely manner. Resorts and casinos, for example, have only a short window of opportunity to turn around a customer experience that's going south fast. Big data analytics gives these businesses the ability to collect customer data, apply analytics and immediately identify potential problems before it's too late.

Government

Certain government agencies face a big challenge: tighten the budget without compromising quality or productivity. This is particularly troublesome with law enforcement agencies, which are struggling to keep crime rates down with relatively scarce resources. And that's why many agencies use big data analytics; the technology streamlines operations while giving the agency a more holistic view of criminal activity.

Health Care

Big data is a given in the health care industry. Patient records, health plans, insurance information and other types of information can be difficult to manage – but are full of key insights once analytics are applied. That's why big data analytics technology is so important to heath care. By analyzing large amounts

of information – both structured and unstructured – quickly, health care providers can provide lifesaving diagnoses or treatment options almost immediately.

Retails

Customer service has evolved in the past several years, as savvier shoppers expect retailers to understand exactly what they need, when they need it. Big data analytics technology helps retailers meet those demands. Armed with endless amounts of data from customer loyalty programs, buying habits and other sources, retailers not only have an in-depth understanding of their customers, they can also predict trends, recommend new products – and boost profitability.

HOW IT WORKS AND KEY TECHNOLOGIES?

Data Management

Data needs to be high quality and well-governed before it can be reliably analyzed. With data constantly flowing in and out of an organization, it's important to establish repeatable processes to build and maintain standards for data quality. Once data is reliable, organizations should establish a master data management program that gets the entire enterprise on the same page.

Data Mining

Data mining technology helps you examine large amounts of data to discover patterns in the data – and this information can be used for further analysis to help answer complex business questions. With data mining software, you can sift through all the chaotic and repetitive noise in data, pinpoint what's relevant, use that information to assess likely outcomes, and then accelerate the pace of making informed decisions.

Hadoop

This open source software framework can store large amounts of data and run applications on clusters of commodity hardware. It has become a key technology to doing business due to the constant increase of data volumes and varieties, and its distributed computing model processes big data fast. An additional benefit is that Hadoop's open source framework is free and uses commodity hardware to store large quantities of data.

In-memory Analytics

By analyzing data from system memory (instead of from your hard disk drive), you can derive immediate insights from your data and act on them quickly. This technology is able to remove data prep and analytical processing latencies to test new scenarios and create models; it's not only an easy way for organizations to stay agile and make better business decisions, it also enables them to run iterative and interactive analytics scenarios.

Predictive Analytics

Predictive analytics technology uses data, statistical algorithms and machine-learning techniques to identify the likelihood of future outcomes based on historical data. It's all about providing a best assessment on what will happen in the future, so organizations can feel more confident that they're making the best possible business decision. Some of the most common applications of predictive analytics include fraud detection, risk, operations and marketing.

Text mining

With text mining technology, you can analyze text data from the web, comment fields, books and other text-based sources to uncover insights you hadn't noticed before. Text mining uses machine learning or natural language processing technology to comb through documents – emails, blogs, Twitter feeds, surveys, competitive intelligence and more – to help you analyze large amounts of information and discover new topics and term relationships.

Benefits of Big Data Analytics

Enterprises are increasingly looking to find actionable insights into their data. Many big data projects originate from the need to answer specific business questions. With the right big data analytics platforms in place, an enterprise can boost sales, increase efficiency, and improve operations, customer service and risk management.

Webopedia parent company, QuinStreet, surveyed 540 enterprise decision-makers involved in big data purchases to learn which business areas companies plan to use Big Data analytics to improve operations. About half of all respondents said they were applying big data analytics to improve customer retention, help with product development and gain a competitive advantage.

Notably, the business area getting the most attention relates to increasing efficiency and optimizing operations. Specifically, 62 percent of respondents said that they use big data analytics to improve speed and reduce complexity.

HOW TO BECOME A BIG DATA ANALYST?

1) Programming

While traditional data analyst might be able to get away without being a full-fledged programmer, a big data analyst needs to be very comfortable with coding. One of the main reasons for this requirement is that big data is still in an evolution phase. Not many standard processes are set around the large complex datasets a big data analyst has to deal with. A lot of customization is required on daily basis to deal with the unstructured data. Which languages are required – R, Python, Java, C++, Ruby, SQL, Hive, SAS, SPSS, MATLAB, Weka, Julia, Scala. As you can not knowing a language should not be a barrier for a big data scientist. At the minimum one needs to know R, Python, and Java. While working you may end up using various tools. Programming Language is only a tool and more tools you have in your kitty, merrier it is.

2) Data Warehousing

Experience with relational and non -relational database systems is a must. Examples of non-relational database include – Mysql, Oracle, DB2. Examples of non-relational database include – NoSql : Hbase, HDFS, MongoDB, CouchDB, Cassandra, Teradeta, etc.

3) Computational frameworks

A good understanding and familiarity with frameworks such as Apache Spark, Apache Storm, Apache Samza, Apache Flink and the classic MapReduce and Hadoop. These technologies help in Big Data processing which can be streamed to a great extent.

4) Quantitative Aptitude and Statistics

While the processing of Big Data requires great use of technology, fundamental to any analysis of data is good knowledge of Statistics and linear algebra. Statistics is a basic building block of data science and understanding of core concepts like summary statistics, probability distribution, random variables, Hypothesis testing framework is important if you are data scientist of any genre.

5) Business Knowledge

To keep the analysis focused, to validate, sort, relate, evaluate the data, the most critical skill of a big data scientist is to have a good knowledge of the domain one is working on. In fact, the reason big data analysts are so much in demand is that its very rare to find resources who have a thorough understanding of technical aspects, statistics and business. There are analysts good in business and statistics but not in programming. There are expert programmers without the knowhow of how to put the programs in the context of the business goal.

CONCLUSION

The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produced a unique moment in the history of data analysis. The convergence of these trends means that we have the capabilities required to analyze astonishing data sets quickly and cost-effectively for the first time in history. These capabilities are neither theoretical nor trivial. They represent a genuine leap forward and a clear opportunity to realize enormous gains in terms of efficiency, productivity, revenue, and profitability.

The Age of Big Data is here, and these are truly revolutionary times if both business and technology professionals continue to work together and deliver on the promise.

REFERENCES

1. Worldometers, “Real time world statistics,” 2014, http://www .worldometers.info/world-population/
2. D. Che, M. Safran, and Z. Peng, “From Big Data to Big Data Mining: challenges, issues, and opportunities,” in Database Systems for Advanced Applications, pp. 1–15, Springer, Berlin, Germany, 2013.
3. M. Chen, S. Mao, and Y. Liu, “Big data: a survey,” Mobile Networks and Applications, vol. 19, no. 2, pp. 171–209, 2014.
4. S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, “Big data: issues and challenges moving forward,” in Proceedings of the IEEE 46th Annual Hawaii International Conference on System Sciences (HICSS ’13), pp. 995–1004, January 2013.
5. R. Cumbley and P. Church, “Is “Big Data” creepy?” Computer Law and Security Review, vol. 29, no. 5, pp. 601–609, 2013.
6. S. Hendrickson, Getting Started with Hadoop with Amazon’s Elastic MapReduce, EMR, 2010.
7. M. Hilbert and P. Lopez, “The world’s technological capacity ´ to store, communicate, and compute information,” Science, vol. 332, no. 6025, pp. 60–65, 2011.