IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Survey on Big Data Analytics

[1]Sneha N P , [2]Keerthana M M

[1] Assistant Professor, [2] Assistant Professor

[1]Department of Computer Science and Engineering,

[1]ATME COLLEGE OF ENGINEERING ,Mysuru, India

Abstract

Enterprises determined to retain their leadership positions in tomorrow's world by moving analytics from the traditional to the Big Data paradigm. But Big Data systems that have poorly integrated business intelligence frameworks, weak data scrubbing rules or non-real-time data processing are unlikely to be fail-safe – with cost prohibitive consequences. Big Data errors can also surface repeatedly unless Assurance teams implement full proof data validation. This paper proposes an approach for Big Data frameworks that avoids pitfalls and delivers the overall process of extracting insights from big data, different types of data, business benefits of big data. This paper provides some understandings on what big data analytics are, how they work, scope, benefits, some of their applications.

Keywords: Big data, Big data analytics, Hadoop

Introduction

Big data is a relative term describing a situation where the volume, velocity and variety of data exceed an organization's storage or compute capacity for accurate and timely decision making. In other words, "Big Data" is a set of data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it. These collection of data is so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include cost, capture, storage, search, sharing, transfer, analysis, visualization and return-on-investment.

Many of the important big data sets encountered in practice assume to be form of a record of transactions. Companies are combining marketing, sales, customer data, transactional data, social conversations and even external data like stock prices, weather and news to identify correlation and causation statistically valid models to help them make more accurate decisions.

Data sets are growing in size because they are being gathered by ubiquitous information hungry mobile devices, global social media sites, increasingly sophisticated consumer facing company websites and swathes of Business to Business (B2B) and Business to Consumer (B2C) Business Intelligence (BI) processes. However, Big Data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make your business more agile and to answer questions that were previously considered beyond your reach. Until now, there was no practical way to harvest this opportunity.

Big Data Analytics

- Examining large amount of data
- Appropriate information
- Identification of hidden patterns, unknown correlations
- Competitive advantage
- Better business decisions: strategic and operational
- Effective marketing, customer satisfaction, increased revenue

Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits[1].

The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modelers and other analytics professionals to analyze large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence(BI) programs[2]. That could include Web server logs and Internet clickstream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet of Things (IoT). In some cases, Hadoop clusters and NoSQL systems are being used as landing pads and staging areas for data before it gets loaded into a data warehouse for analysis, often in a summarized form that is more conducive to relational structures. Increasingly though, big data vendors are pushing the concept of a Hadoop data lake that serves as the central repository for an organization's incoming streams of raw data. In such architectures, subsets of the data can then be filtered for analysis in data warehouses and analytical databases, or it can be analyzed directly in Hadoop using batch query tools, stream processing software and SQL on Hadoop technologies that run interactive, ad hoc queries written in SQL.

Potential pitfalls that can trip up organizations on big data analytics initiatives include a lack of internal analytics skills and the high cost of hiring experienced analytics professionals. The amount of information that's typically involved, and its variety, can also cause data management headaches, including data quality and consistency issues. In addition, integrating Hadoop systems and data warehouses can be a challenge, although various vendors now offer software connectors between Hadoop and relational databases, as well as other data integration tools with big data capabilities. Big Data is Usually defined by 4 V's i.e. Velocity, Volume, Variety and Value as shown in the Figure 1.1

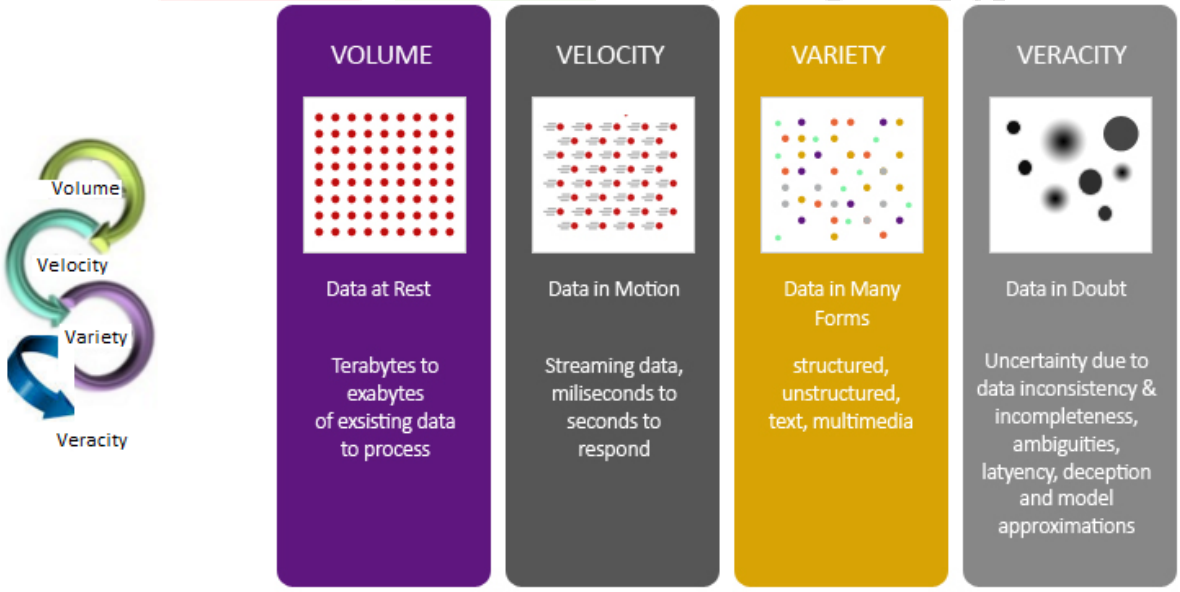

Figure 1.1 Big Data is Usually defined by 4 V's

The overall process of extracting insights from big data can be broken down into five stages These five stages form the two main sub-processes: data management and analytics. Data management involves processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis. Analytics, on the other hand, refers to techniques used to analyze and acquire intelligence from big data. Thus, big data analytics can be viewed as a sub-process in the overall process of 'insight extraction' from big data as shown in the figure 1.2.

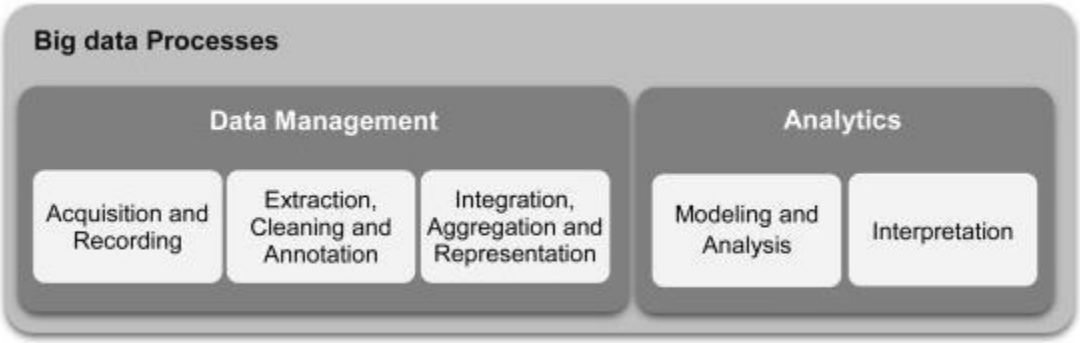

Figure 1.2 Big data process

Big Data generates value from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques[3]. For Example, Decoding the human genome originally took l0years to process[5]; now it can be achieved in one week as shown in the figure 1.3.

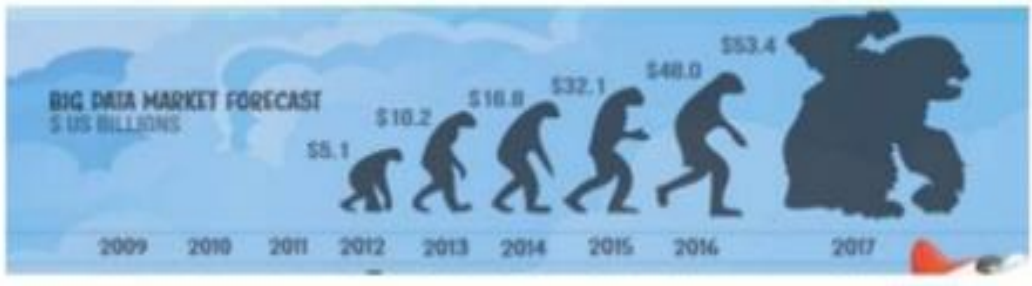

Figure 1.3 human genome

Below are the different types of data:

- Relational Data (Tables/Transaction/Legacy Data) and Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
- Social Network, Semantic Web (RDF), …
- Streaming Data
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc.…

Big data is difficult to work with using most relational database management systems and desktop statistical software and visualization packages as it requires huge amounts of parallel software running on tens, hundreds, or even thousands of servers.

Business Benefits of Big Data

The benefits of Big Data analytics come from building business capabilities that enable a company to gain competitive advantage in increasing their market share. These capabilities need to be built along multiple dimensions, namely process, technology, and people. By not considering these dimensions while planning Big Data investments, companies fail to derive the desired value from their investment. For instance, one may have a Big Data technology setup; however, if the processes are not capturing the required data, then the analytics are only going to be as good (or as bad) as the underlying data. Another possibility is that one may end up with volumes of data, but the business users do not know what to do with it or how to interpret the results.

There are compelling business reasons for developing Big Data analytical capabilities:

Performance Management: Performance management involves understanding the meaning of data in company databases using pre-determined queries and multidimensional analysis. The data used for this analysis are transactional, for example, years of customer purchasing activity or inventory levels and turnover. Managers can ask questions such as which are the most profitable customer segments and get answers in real-time that can be used to help make short-term business decisions and longer term plans. The main challenge is to ensure the quality and completeness of transactions entered into the system or the result will be "garbage in, garbage out." Also, to guarantee a complete picture of the business, multiple databases across functions have to be integrated.

Data Exploration: Data exploration makes heavy use of statistics to experiment and get answers to questions that managers might not have thought of previously. This approach leverages predictive modelling techniques to predict user behavior based on their previous business transactions and preferences. Cluster analysis can be used to segment customers into groups based on similar attributes. Once these groups are discovered, managers can perform targeted actions such as customizing marketing messages, upgrading services and cross/ up-selling to each unique group. Another popular use case is to predict what group of users may "drop out." Armed with this information, managers can proactively devise strategies to retain this user segment and lower the churn rate.

Social Analytics: Social analytics measure the vast amount of non-transactional data that exists today. Much of this data exist on social media platforms, such as conversations and reviews on Facebook, Twitter, and Google+. Social analytics measure three broad categories: awareness, engagement, and word-of-mouth or reach.

Decision Science: Decision science involves experiments and analysis of non-transactional data, such as consumer-generated product ideas and product reviews, to improve the decision-making process. Unlike social analysers who focus on social analytics to measure known objectives, decision scientists explore social big data as a way to conduct "field research" and to test hypotheses. Crowdsourcing, including idea generation and polling, enables companies to pose questions to the community about its products and brands. Decision scientists, determine the value, validity, feasibility and fit of these ideas and eventually report on if/ how they plan to put these ideas in to action.

With respect to **future trends in the Big Data field**, the following practice is starting to emerge:

- Integrating multiple big data strategies: While a company can be effective with a single Big Data strategy, the most effective companies leveraging Big Data today are combining strategies. For example, one financial institution is leveraging both Social Analytics (non-transactional social data) and Performance Management (business intelligence using transactional data) strategies to guide its customer service. The bank traditionally determined its "top" customers based on metrics such as number and balance of accounts; these were the customers who received premium service. Now, the bank is planning to incorporate social metrics into the equation. Those online customers who are very active with respect to mentioning, engaging with, and promoting the bank on social channels will also be considered for high-level service programs.
- Re-develop your products**:** Big Data can also help you understand how others perceive your products so that you can adapt them, or your marketing, if need be. Analysis of unstructured social media text allows you to uncover the sentiments of your customers and even segment those in different geographical locations or among different demographic groups. On top of that, Big Data lets you test thousands of different variations of computer-aided designs in the blink of an eye so that you can check how minor changes in, for instance, material affect costs, lead times and performance.
- Perform risk analysis: Success not only depends on how you run your company. Social and economic factors are crucial for your accomplishments as well. Predictive analytics, fueled by Big Data allows you

to scan and analyze newspaper reports or social media feeds so that you permanently keep up to speed on the latest developments in your industry and its environment.

- Keeping your data safe: You can map the entire data landscape across your company with Big Data tools, thus allowing you to analyses the threats that you face internally. You will be able to detect potentially sensitive information that is not protected in an appropriate manner and make sure it is stored according to regulatory requirements.
- Create new revenue streams: The insights that you gain from analyzing your market and its consumers with Big Data are not just valuable to you. We could sell them as non-personalized trend data to large industry operating in the same segment and create a whole new revenue stream. One of the more impressive examples comes from Shazam, the song identification application. It helps record labels find out where music sub-cultures are arising by monitoring the use of its service, including the location data that mobile devices so conveniently provide. The record labels can then find and sign up promising new artists or remarket their existing ones accordingly.
- Customize your website in real time: Big Data analytics allows you to personalize the content or look and feel of your website in real time to suit each consumer entering your website, depending on, for instance, their sex, nationality or from where they ended up on your site. The best-known example is probably offering tailored recommendations: Amazon's use of real-time, item-based, collaborative filtering (IBCF) to fuel its 'Frequently bought together' and 'Customers who bought this item also bought' features or LinkedIn suggesting 'People you may know' or 'Companies you may want to follow'. And the approach works: Amazon generates about 20% more revenue via this method.
- Offering enterprise-wide insights: Previously, if business users needed to analysis of large amounts of varied data, they had to ask their IT colleagues for help as they themselves lacked the technical skills for doing so. Often, by the time they received the requested information, it was no longer useful or even correct. With Big Data tools[4], the technical teams can do the groundwork and then build repeatability into algorithms for faster searches. In other words, we can develop systems and install interactive and dynamic visualization tools that allow business users to analysis, view and benefit from the data.
- Real-time Fraud Detection: Integrated more than 60 different data sources, analyzed 60 million data sets/day.

Conclusion

Analytics for the 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 analyzed, 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 analyzing the data, from acquiring the data and discovering new insights to making repeatable decisions and scaling the associated information systems.

References

1. Hu Shuijing,' Big Data Analytics: Key Technologies and Challenges', International Conference on Robots & Intelligent System (ICRIS), 10.1109/ICRIS.2016.30,2016
2. Anku Jaiswal, Purrushottam Bagale,' A Survey on Big Data in Financial Sector', International Conference on Networking and Network Applications, IEEE DOI 10.1109/NaNA,2017
3. Sayali Gaikwad, Pranali Nale, Ravindra Bachate,'Survey on Big data Analytics for digital world', 2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology, 10.1109/ICAECCT.2016.7942579, 2016
4. J. Vijayaraj, R. Saravanan, P. Victer Paul,'A comprehensive survey on big data analytics tools', International Conference on Green Engineering and Technologies (IC-GET), 10.1109/GET.2016.7916733,2016
5. D K Thara, B G Premasudha, V Ravi Ram,'Impact of big data in healthcare: A survey', 2nd International Conference on Contemporary Computing and Informatics, 10.1109/IC3I.2016.7918057,2017.