Asian Journal of Applied Science and Technology (AJAST)
Volume 1, Issue 1, Pages 233-234, February 2017



# A Review on Big Data Analytics

Dr.R.Reka[1], K.Saraswathi[2] and Dr.K.Sujatha[3]

[1]*Professor, Department of Information Technology, Panimalar Institute of Technology, Chennai, India. Email: rekasatheesh@gmail.com*
[2]*Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy, India. Email: saraswathimuruganmsc@gmail.com*
[3]*Associate Professor, Department of Computer Science and Engineering, Panimalar Institute of Technology, Chennai, India. Email: sujatha.ssps@gmail.com*

*Article Received: 17 February 2017* *Article Accepted: 27 February 2017* *Article Published: 28 February 2017*

ABSTRACT

Big data is a data or data sets so large or complex that traditional data processing applications are inadequate and distributed databases are needed. Firms like Google, eBay, LinkedIn, and Face book were built around big data from the beginning. It is a collection of massive and complex data sets that include the huge quantities of data, social media analytics, data management capabilities, real time data etc. Challenges include sensor design, capture, sharing, storage, analysis, visualization, information privacy etc. Big data refers to datasets high in variety and velocity. Big Data Analytics refers to techniques used to analyze and acquire intelligence from big data [1].Thus, big data analytics can be viewed as a sub-process in the overall process of 'insight extraction' from big data.

Keywords: Big Data, Big Data Analytics, Text, Image, Audio, Video, Web, Multimedia, Mobile and Social Media.

## 1. INTRODUCTION

This paper is organized as follows. We begin the paper by defining big data. We highlight the fact that size is only one of several dimensions of big data. Other characteristics, such as the frequency with which data are generated, are equally important in defining big data. We then expand the discussion on various types of big data, namely text, audio, video, web, multimedia, mobile and social media [2].

### *1.1 Types of Big Data*

#### *A. Text Analytics*

Text analytics refers to the extraction of useful information from text sources. It is a broad term that describes tasks from annotating text sources with meta-information such as people and places mentioned in the text to a wide range of models about the documents (e.g., sentiment analysis, text clustering, and categorization)[3]. Text analytics (text mining) refers to techniques that extract information from textual data. Social network feeds, emails, blogs, online forums, survey responses, corporate documents, news, and call center logs are examples of textual data held by organizations. Text summarization techniques automatically produce a succinct summary of a single or multiple documents. The resulting summary conveys the key information in the original text(s). Applications include scientific and news articles, advertisements, emails, and blogs. Broadly speaking, summarization follows two approaches: the extractive approach and the abstractive approach. In extractive summarization, a summary is created from the original text units (usually sentences). The resulting summary is a subset of the original document. In contrast, abstractive summarization techniques involve extracting semantic information from the text. The summaries contain text units that are not necessarily present in the original text. Question answering (*QA*) techniques provide answers to questions posed in natural language. Apple's Siri and IBM's Watson are examples of commercial QA systems. These systems have been implemented in healthcare, finance, marketing, and education [1].Sentiment analysis *(*opinion mining*)* techniques analyze opinionated text, which contains people's opinions toward entities such as products, organizations, individuals, and events. Businesses are increasingly capturing more data about their customers' sentiments that has led to the proliferation of sentiment analysis (Liu, 2012). Marketing, finance, and the political and social sciences are the major application areas of sentiment analysis. Sentiment analysis techniques are further divided into three sub-groups, namely document-level, sentence-level, and aspect-based. Document-level techniques determine whether the whole document expresses a negative or a positive sentiment. Sentence-level techniques attempt to determine the polarity of a single sentiment about a known entity expressed in a single sentence.

#### *B. Image Data Analytics*

Image analytics is the automatic algorithmic extraction and logical analysis of information found in image data using digital image processing techniques. The use of bar codes and QR codes are simple examples, but interesting examples are as complex as facial recognition and position and movement analysis. Today, images and image sequences (videos) make up about 80 percent of all corporate and public unstructured big data. As growth of unstructured data increases, analytical systems must assimilate and interpret images and videos as well as they interpret structured data such as text and numbers. An image is a set of signals sensed by the human eye and processed by the visual cortex in the brain creating a vivid experience of a scene that is instantly associated with concepts and objects previously perceived and recorded in one's memory. To a computer, images are either a raster image or a vector image. Simply put, raster images are a sequence of pixels with discreet numerical values for color; vector images are a set of color-annotated polygons. To perform analytics on images or videos, the geometric

 

encoding must be transformed into constructs depicting physical features, objects and movement represented by the image or video. These constructs can then be logically analyzed by a computer.

The process of transforming big data (including image data) into higher-level constructs that can be analyzed is organized in progressive steps that each adds value to the original information in a value chain (see Figure 2) – a concept developed by Harvard professor Michael Porter. Prescriptive analytics leverages the emergence of big data and computational and scientific advances in the fields of statistics, mathematics, operations research, business rules and machine learning [4].

### *C. Audio Data Analytics*

Audio analytics is the process of compressing data and packaging the data in to single format called audio. Audio Analytics refers to the extraction of meaning and information from audio signals for Analysis. There are two way to represent the audio Analytics is 1) Sound Representation 2) Raw Sound Files. Audio file format is a format for store digital audio data on a system. There are three main audio formats: Uncompressed audio format, Lossless compressed audio format, Lossy compressed audio format.

### *D. Video Data Analytics*

Video is a major issue when considering big data. Videos and images contribute to 80 % of unstructured data. Now a day, CCTV cameras are the one form of digital information and surveillance. All these information is stored and processed for further use, but video contains lots of information and is generally large in size. For example YouTube has innumerable videos being uploaded every minute containing massive information. Not all video are important and viewed largely. This creates a situation where videos create a junk and hard-core contribution to big data problems. Apart from videos, surveillance cameras generate a lot of information in seconds. Even a small Digital camera capturing an image stores millions of pixel information in mille seconds [5].

### *E. Web Analytics*

The aim of Web analytics is to retrieve, extract the information from Web Pages. Web Analytics also called Web mining.

### *F. Multimedia Analytics*

Recently multimedia data, including images, audio, and video has grown at a tremendous rate. Multimedia analytics refers to extract interesting knowledge and semantics captured in multimedia data. Multimedia analytics covers many subjects like Audio Summarization, Multimedia annotation, Multimedia indexing and retrieval.

### *G. Mobile Analytics*

Mobile data traffic increased 885PBs Per Month at the end of 2012. Vast volume of application and data leads to mobile analytics. Mobile analytics involves RFID, mobile phones, Sensors etc.

### *H. Social Media*

The Social Media analytics is collecting information or data from the social media websites, blogs etc. and uses it in business purpose or decision making. Now a Day Social Media is the best platform for understand the real-time customer choice or intentions and sentiments, using social media business advertising, product marketing easily. EBay.com uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising. Inside eBay's 90PB data warehouse. Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB. Facebook handles 50 billion photos from its user base. As of August 2012, Google was handling roughly 100 billion searches per month [6].

## 2. CONCLUSION

Now, computer industry accepts Big Data as a new challenge for all types of machine automated systems. There are many issues in storage, management, and retrieval of data known as Big Data. The main problem is how we can use this data for increasing business and improvement in living standard of people. In this paper we are discussing the Big Data Analytics. It will motivate researchers for finding knowledge from the big amount of data available in different forms in different areas.

## REFERENCES

[1] Mereena Thomas, "A Review Paper on Big Data", International Research Journal of Engineering and Technology, Volume: 02 Issue: 09 | Dec-2015.

[2] AmirGandomi, Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management,* Volume 35, Issue 2, April 2015, Pages 137-144.

[3] Niels Kasch, Mariann Micsinai, "Text Analytics and Natural Language Processing in the Era of Big Data", *Pivotal,* October 23, 2014.

[4] Fritz Venter, Andrew Stein, "Sizing up the potential impact of prescriptive analytics driven by Proliferation of images and video", *Images & videos: really big data* November 2012.

[5] Jai Prakash Verma, Smita Agrawal, Bankim Patel**,** Atul Patel, "Big Data Analytics: Challenges and Applications for Text, Audio, Video, and Social Media Data", International Journal on Soft Computing, *Artificial Intelligence and Applications,* Vol.5, No.1, February 2016.

[6] Ankita S. Tiwarkhede, Prof. Vinit Kakde, "A Review Paper on Big Data Analytics*", International Journal of Science and Research,* Volume 4 Issue 4, April 2015.

 