BIG DATA AT A GLANCE FROM A LAYMAN'S PERSPECTIVE

Ali Akbar Petiwala *

*Librarian,

RBK School,

Mumbai, Maharashtra,

India.

QR Code

ABSTRACT: -

In the world of IT or if you are a Librarian, you have doubtlessly heard about the Big Data phenomenon. You might have even encountered professionals who introduced themselves as data scientists. Hence, you are wondering, just what is this emerging new area of science? What types of knowledge and problem-solving skills do data scientists have? What types of problems are solved by data scientists through Big Data tech? This paper will help user to prepare for a deeper understanding of the important areas of Big Data and other areas of it.Big data is new to many people, so it requires some investigation and understanding. Many different people need knowledge about big data. Paper explains the basic concepts we need for a full understanding of big data, from both a layman's and a business perspective. Paper also introduce few major concepts and components so that in a way it is meaningful conversing about big data.

KEYWORDS: Big Data, Structure Data, Unstructured Data, Cloud-based

INTRODUCTION

Every day, our society is creating about 3 quintillion bytes of data every day. You are probably wondering what 3 quintillion is. Well, this is 3 followed by 18 zeros. With this massive stream of data, the need to make sense for Big Data understanding is now a crucial point. As everyone is uploading some random data every day we must have basic knowledge of big data.

In this digital era, where everything and anything we do leaves a digital trace we call data. At the center of this digital universe is the World Wide Web from which comes a torrent of data that floods our notice every single second. With well over one trillion web pages (50 billion of which have already been indexed by and are searchable through various major search engines), the web offers us unparalleled interconnectivity which allows us to interact with anyone and anything within a connected network we happen to be part of. As a consequence, the web is continuously overflowing with massive data so huge that it is

almost impossible to digest or crunch into usable segments for practical applications – if they are of any use at all. This enormous, ever growing data that goes through and are stored in the web together with the developing technologies designed to handle it is what is collectively referred to as 'big data'. 'Big data' is one of the latest technology trends that is extremely affecting the way organizations utilize information to enhance the customer experience, improve their products and services and even efficiently manage health care services.

SO, WHAT DOES BIG DATA LOOK LIKE?

If you want to have an idea on what 'big data' really looks like so think of the web which currently covers more than 100 million domains and is still growing at the rate of 20,000 new domains every single day. The data that comes from these domains is so massive that it is practically immeasurable much less manageable by any conventional data management and retrieval methods that are available today. And that is only for our starters. Add to this the 300 million daily Facebook posts, 60 million daily Facebook updates, and 250 million daily tweets coming from more than 900 million combined Facebook and Tweeter users. Moreover data coming from over six billion smart phones currently in use today which continually access the internet to do business online, to post status updates on social media, send out tweets,GPS and

many other digital transactions. Consider more than 10,000 credit card payments being done online or through other connected devices every single second.

The 'big data' that had been accumulated by the web for the past five years (since 2010) and are now stored in millions of servers scattered all over the globe. The 'big data' we refer to includes anything and everything that has been fed into big data systems such as social network chatters, content of web pages, GPS trails, financial market data, online banking transactions, streaming music and videos, podcasts, satellite imagery, etc. It is estimated that over 2.5 quintillion bytes of data (2.5×10^{18}) is created by us every day. This massive flood of data which we collectively call as 'big data' just keeps on getting bigger and bigger through time. Experts estimate that its volume will reach 35 zetta bytes (35×10^{21}) by 2020. When this data sets grow extremely big or become excessively too complex for traditional data management tools to handle, it is considered as 'Big Data'.
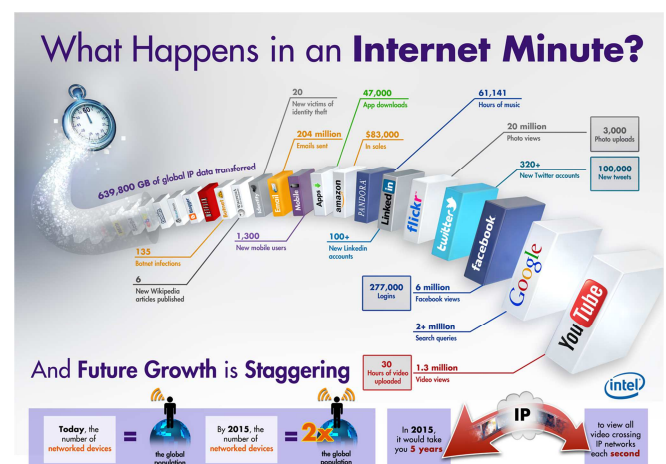

Figure 1: Infographic from Intel.com

UNDERSTANDING BIG DATA BETTER

Big data is not a single entity. Rather, it is a mixture of several data-management technologies that have evolved over time. Big data is what allows businesses the ability to store, analyze, and exploit massive amounts of data with great ease and on real time to gain deeper market insights and create new value that will benefit the organization. But, big data has to be managed well before it can be utilized and the key to managing big data well is by having a clear understanding of what it truly is. Unfortunately, each time we attempt to define the term big data, our minds almost always end up swirling in confusion. The data sets that make up big data are varied and include both structured and unstructured data. In essence, big data is a mixture of unstructured and multi-structured data which together compose the bulk of information contained therein. This varied customer data includes information coming from the Customer Relations Management systems; feedbacks, reactions, and interactions from social media; call-center logs, etc. they are part and parcel of big data analytics and contribute a lot to customer.
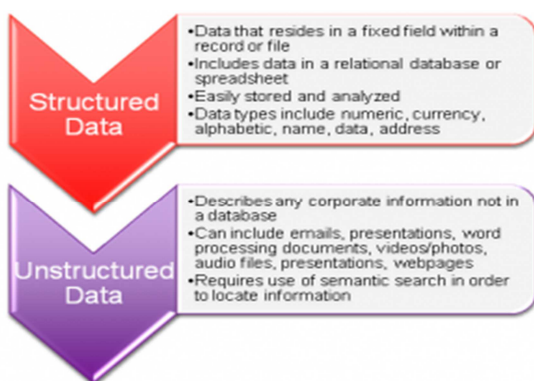

Figure 2: Data Ladder
(Structure & Unstructure Data)

Structured data refers to any data that are seamlessly contained in relational databases and spreadsheets. It is so well organized that that they are easily searchable by even the simplest search engine algorithm. It can also be entered, stored, analyzed and queried in relational databases with great ease. Samples of structured data include numeric, currency, alphabetic, name, data, address, etc.

Unstructured data refers to data sets that are text-heavy and are not organized into specific fields. Because of this, traditional databases or data models have difficulty interpreting them. Examples of unstructured data include Metadata, photos and graphic images, webpages, PDF files, wikis and word processing documents, streaming data, blog entries, videos, emails, Twitter tweets, and other social media posts.

BIG DATA AND THE FUTURE

The ability to store data at a fast rate is not entirely new. The area that is new is the ability to make something useful with this data with cost-efficiency in mind. For decades, governments and multi-national companies have been storing and processing tons of data. Aside from the new capacities in handling huge amounts of data, we are also observing the rise of new technologies that are created to manage complex, non-conventional data - specifically the types of semi-structured or unstructured data produced by web logs, sensors, census reports, warranties, customer

service records, mobile networks, and social media. Previously, data should be neatly organized in sets or tables. The capacity to handle large and complicated sets of data has not reduced the demand for more size and faster speed. Each day, it seems that a new technology app is launched, which drives the Big Data technology further.

Big Data really matters for big enterprises like Amazon, Google, and Facebook and there is no question about it. They have been using it all along, albeit successfully. To them, big data analytics has become not only part and parcel of their marketing and operational mix but now serves as one of the pillars of their foundation. They rely heavily on it to come up with creative and innovative ideas to serve their customers better. Suppose when Amazon started capturing and collecting massive customer data and stored all the searches and all the transactions made by their customers along with every single piece of information available, the same customers information was available in a segregated form on cloud. This information is made available in a shared form with the other vendors. Big Data may grow too big too soon such that we may find it difficult to create new value from it. This may put us in a situation where we may have to choose the quality of the data over the quantity.

BIG DATA CHALLENGES

IT leaders are still struggling to understand the concept of big data. There is no way they can develop and implement a workable big data strategy this way. They have to understand the concept and hurdle the key big data challenges before they can bring their organization to level up and put a big data system in place and working perfectly for them. Many of the business and IT leaders are still unclear about the concept of big data and are still struggling to understand the many benefits it had to offer their respective businesses. We cannot expect all business and IT leaders to be familiar with much less understand these innovative big data solutions immediately. They will have to learn the latest technologies to keep abreast of the latest developments before they can utilize big data analytics more effectively to their company's advantage. And, learning must be on a continuing basis. They may even have to deal with different big data solutions providers and partners in bigger numbers than those they have dealt with in the past. A the same time, they have to balance the costs of subscribing to big data solutions provider with their actual business needs by limiting their service subscriptions only to what they need at the moment relative to the project at hand. Cloud based big data solutions offer huge tremendous cost saving opportunities and storage facility. With cloud based big data solutions company data can be stored and managed in various data centers remotely located all around the globe.

CONCLUSION

Big data is not a single technology but a combination of old and new technologies that helps companies gain actionable understanding. Therefore, big data is the capability to manage a huge volume of data, at the right speed, and within the right time frame to allow real-time analysis and reaction.

- ❖ Volume: How much data
- ❖ Velocity: How fast that data is processed
- ❖ Variety: The various types of data

The paper concludes with overview of big data and introduced all its significant components. The paper has also enlightened users about the data structures that are foundational to big data.

REFRENCES

•Hurwitz, J. (2008). Big data for dummies. Hoboken, NJ: Wiley.

•Mosco, V. (2014). To the cloud: Big data in a turbulent world. Routledge.

•Sakr, S., & Gaber, M. M. (2014). Large scale and big data: Processing and management.

•Yu, S., & Lin, X. (2015). Networking for big data.