

Big Data by 3V's and Its Importance

Sameet Sonawane¹, Dhairya Patel², Mansi Kevadiya³, Raj Modi⁴, Jigar Moradiya⁵, Akash Thomas⁶
^{1,2,3,4,5,6}Associate Developer, SRKAY Consulting Group, Surat, India

Abstract: The aim of this paper is to study research about Big data. The term, Big Data 'has been coined to refer to the gargantuan bulk of data that cannot be dealt with by traditional data-handling techniques so, the question that arises now is, how to develop a high performance platform to efficiently analyse big data and how to design an appropriate mining algorithm to find the useful things from big data. Complete structure of big data by 3 V's is going to cover in this paper.

Keywords: Big data, Mining algorithm, Big Data, 3 V's.

1. Introduction

Big data is an evolving term that describes any voluminous amount of structured (database with rows and columns), semi structured (XML and JSON documents) and unstructured data (Photographs and video, Social media data) that has the potential to be mined for information [1]. Big data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. While the problem of working with data that exceeds the computing power or storage of a single computer is not new, the pervasiveness, scale, and value of this type of computing has greatly expanded in recent years [2]. In this article, we will talk about big data on a fundamental level and define common concepts you might come across while researching the subject. There are the three V's of big data: volume, velocity and variety.

2. Aim of the study

The aim of this paper is to study research related to big data and its importance in different sectors.

3. V's of Big Data

The definition of 3Vs implies that the data size is large, the data will be created rapidly, and the data will be existed in multiple types and captured from different sources, respectively. Later studies pointed out that the definition of 3V's is insufficient to explain the big data we face now. Thus, veracity, validity, value, venue, variability, vocabulary, and vagueness were added to make some complement explanation of big data [3]. But there is a main 3V's for a big data.

A. Volume

Now, the data storage is not only as a text data. We can find data in the format of videos, music and images on the social media channels. It is very common to have a storage system of

Terabytes and Petabytes for enterprises. As the database grows the applications and architecture built to support the data needs to be reevaluated quite often. Sometimes the same data is re-evaluated with multiple angles and even though the original data is the same the new found intelligence creates explosion of the data. The big volume indeed represents Transaction-based data stored through the years. Unstructured data streaming in from social media and Satellites as well as increasing amounts of sensor and machine-to-machine data being collected. In the past, too much data volume was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use analytics to create value from relevant data. Fig. 1.

B. Velocity

Data is streaming in at unprecedented speed and must be dealt with in a timely manner. The data growth and social media explosion have changed how we look at the data. There was a time when we used to believe that data of yesterday is recent. The matter of the fact newspapers is still following that logic. However, news channels and radios have changed how fast we receive the news. Today, people reply on social media to update them with the latest happening. On social media sometimes a few seconds old messages (a tweet, status updates etc.) is not something interests users. They often discard old messages and pay attention to recent updates. The data movement is now almost real time and the update window has reduced to fractions of the seconds. This high velocity data represent RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data velocity is a challenge for most organizations. Fig. 1.



Fig. 1. 3V's factors of Big Data

C. Variety

Data can be stored in multiple format. Structured, numeric data and also as database, csv, excel, access or for the matter of the fact, it can be stored in a simple text file. Sometimes the data is not even in the traditional format as we want, it may be in the form of picture, video, SMS, pdf or something we might have not thought about it. Information created from line-of-business applications. Unstructured text documents, email, audio, video, stock ticker data and financial transactions. Managing, merging, processing and governing different varieties of data is something many organizations still grapple with Fig. 1. For example, It is the need of the organization to arrange it and make it meaningful. It will be easy to do so if we have data in the same format, however it is not the case most of the time. The real world has data in many different formats and that is the challenge we need to overcome with the Big Data. We can consider two other important dimensions for big data that are

D. Variability

In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Is something trending in social media? Daily, seasonal and event-triggered peak data loads can be challenging to manage. Even more so with unstructured data involved.

E. Variability

Today's data comes from multiple sources. And it is still an undertaking to link, match, cleanse and transform data across systems. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages or your data can quickly spiral out of control [17].

4. Importance of big data

The importance of big data does not revolve around how much data a company has but how a company utilizes the collected data. Every company uses data in its own way; the more efficiently a company uses its data, the more potential it has to grow. The company can take data from any source and analyze it to find answers which will enable:

- *Cost Savings:* Some tools of Big Data like Hadoop and Cloud-Based Analytics can bring cost advantages to business when large amounts of data are to be stored and these tools also help in identifying more efficient ways of doing business.

- *Time Reductions:* The high speed of tools like Hadoop and in-memory analytics can easily identify new sources of data which helps businesses analyzing data immediately and make quick decisions based on the learning.
- *New Product Development:* By knowing the trends of customer needs and satisfaction through analytics you can create products according to the wants of customers.
- *Understand the market conditions:* By analyzing big data you can get a better understanding of current market conditions. For example, by analyzing customers' purchasing behaviours, a company can find out the products that are sold the most and produce products according to this trend. By this, it can get ahead of its competitors.
- *Control online reputation:* Big data tools can do sentiment analysis. Therefore, you can get feedback about who is saying what about your company. If you want to monitor and improve the online presence of your business, then, big data tools can help in all this. [5]

5. Conclusion

In this paper concept of big data is defined by 3 V's volume, velocity and verity that is a main dimension for the big data understanding as well as other dimensions. Many tools are available for big data handling and can accomplish a number of functions. Big data is very important for each sectors for different functions like business, online marketing and selling, to find the current market scenarios and market analysis.

References

- [1] Vibhavari Chavan, Rajesh. N. Phursule : Survey Paper On Big Data
- [2] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos: Big data analytics: a survey
- [3] <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>
- [4] <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>
- [5] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos, "Big data analytics: a survey," in Journal of Big Data, 2:21, 2015.
- [6] <http://pig.apache.org>
- [7] <https://www.newgenapps.com/blog/importance-benefits-competitive-advantage-big-data>