

Introduction to Big Data (P1)

Aydan Alyanak
Vodafone Nederland
Simon Carmiggeltstraat 6 1011DJ
Amsterdam, Netherlands
aalyanak@indiana.edu

ABSTRACT

In this paper, I'll be describing what's big data, what makes it interesting and what limitations does big data have.

Keywords

Big Data; What's Big Data; What makes Big Data interesting; limitations of Big Data

1. INTRODUCTION

In the last decades, with the increasing range of technology usage and decreasing cost of technology, revolution of big data has been started. Since the computer and internet technology started to being used, the outcome data is stored with no use. All this data is piled up in data warehouses, and especially in late years, the importance of this data is realized and taking advantage of all this complication has become the first point of achievement [3].

2. WHAT IS BIG DATA?

The concept of "Internet of Things" is first mentioned by Kevin Ashton in 1999 in a presentation he made for PG [2]. Ashton made this point to get attention to the fact that, the machines which use Internet and collect data, still needs human control for operation, collection, clarification and filtration to be brought to the format that people can use.

Structured data is much easier to use in traditional analytics, but it's not enough anymore in today's world. All the data stored in warehouses for decades needs to be taken advantage of, but it's not quite simple as it's seems. In the last 10 years, the idea of big data gained importance and attention, the researches and studies are focused on making use of this enormous amount of advantageous source. There is no clear definition for "Big Data". It is defined based on some characteristics called 3V [8].

Volume: The size of Big Data can be defined by Terabytes or Exabytes [8].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2016 ACM. ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

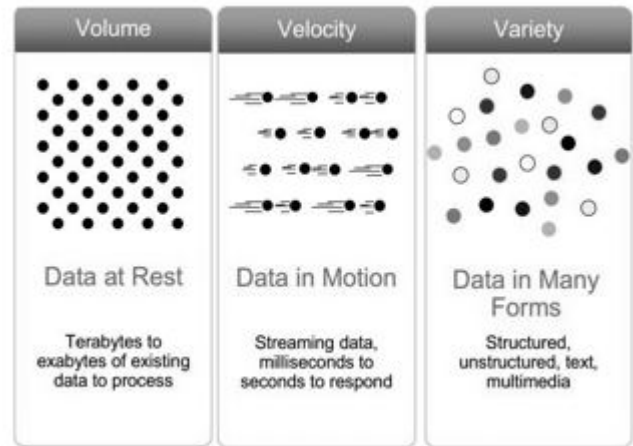


Figure 1: 3V Characteristics of Big Data [1].

Velocity: Even data may be not big or complicated, the velocity of data may increase higher frequencies or periods that brings higher volumes or speed which cause processing become more difficult [1]. Processing frequency may also differ from the user requirements, some data need to be processed real-time and some may only be processed when needed [8].

Variety: The types of big data may differ from source to source. It's nearly impossible to characterize all different types of data in the same format, and it's significantly different from the traditional data [1].

These 3 characteristics are the main points of view for explaining what big data is. Different sources may also mention "Verification/Veracity", "Value" or other characteristics for explaining big data, but the first 3V's have the most common use.

3. WHY IS BIG DATA INTERESTING?

The Data Science certification program of Indiana University brings people all around the world to be a part of real challenge for learning what Big Data is and the wide usage area and different sectors makes it even more interesting.

For me, the customer and campaign analytics sector which I'm currently working in makes Data Science and Big Data very interesting. Everything starts with curiosity, and we're lucky that we started to be interested in this area in the early stages. In my opinion, Big Data will become a very big deal in the coming decade and we'll be a part of it.

Many of the students discussed about what makes big data interesting in discussion board, I could manage them summarize into groups:

Social Media Internet of Things: This is one of the most interesting source of big data. Today's leader online service companies such as Google, Facebook, Twitter, Netflix etc. both produce and store huge amount of data and even started to make use of it. Especially Google is very good at personal assistance in mobile devices [6], it synchronizes with any app in your device and shows you the hours of plane, train or public transport, the maps, tickets, attractions, currencies, weather, and improves itself every day to add new features. It's very amazing.

Application in Different Sectors: In chemistry, health-care, environmental, information technology, marketing, entertainment, online or several other sectors may benefit from big data [4].

The Technology We Can Make Use Of: Especially in software development, ETL tools, big data shows us building new ways out of the old fashioned methods [5]. The satisfaction of user groups or customers lies here, it's important to improve our understanding of big data and learn how to process and make a better use of it.

Many other points can be discussed to explain this incredible area, and it's getting even more interesting when we start to understand, analyse and participate.

4. WHAT LIMITATIONS DOES BIG DATA HAVE?

There are many different studies to explain the limitations of big data analytics, first of all, the greatest limitation lies in the data's name. Data is enormously big and this brings its own limitation. It's important to analyse the data from right perspective, and the size of data limits us to start and continue with a correct point of view [7].

Data is relatively new to us compared with traditional data, and not enough people have the skill set to process. It's still in the potential phase, companies still have the data but don't know how to use.

Data is complicated and difficult to understand. It's not structured and before using it, it's important to bring it to a correct format.

It's also important for us not to overkill the process. It's very open to ask wrong questions, look from incorrect point of views and spend large amount of time or money to the parts we don't even need. The limitation in here is also lying under the lack of knowledge and certainty.

These limitations can be varied more, and the solution is again more education and training on the area. Solution comes with knowledge, and at the end, experience. When people get more experienced in the area, use cases will grow and limitations will fade out quickly.

5. CONCLUSION

It's fascinating that we pay more and more attention to big data every year, and many big and leader companies started to hire more people to focus on big data analytics and data science. It's a huge source we can benefit, but the most important fact is that we need to understand and transform it to a finalized format that we can make use of. It's also important for us not to lose our way in the middle, which is very likely and possible. The big data has its own

limitations, and to avoid from them and succeed at the end comes by knowledge and experience.

6. REFERENCES

- [1] M. Acungil. Where is big data and where did it come from?, June 2014.
- [2] K. Ashton. That 'internet of things' thing in the real world, things matter more than ideas. June 2009.
- [3] Y. Goksen. Big data, 2015.
- [4] Y. Hong. Why is big data interesting to me?, Augu6t 2016.
- [5] V. Navitsky. Why is big data interesting to me?, Augu6t 2016.
- [6] H. Pal. Why is big data interesting to me?, Augu6t 2016.
- [7] N. Schenker. Big data: Uses and limitations, September 2013.
- [8] A. Zaslavsky, C. Perera, and D. Georgakopoulos. Sensing as a service and big data. pages 1–3, July 2012.