COMMUNICATION SKILLS
Master of Science in Computational Mechanics
Master Métodos Numéricos en Ingeniería

BIG DATA - HOW DOES IT AFFECT US?

MARTIN VEE AKSELSEN* AND HENRIK VAN THRIEL†

* Universidad Politécnica de Cataluña
Campus Norte UPC, 08034 Barcelona, Spain
e-mail: martin.vee.akselsen@upc.edu,

†Universidad Politécnica de Cataluña
Campus Norte UPC, 08034 Barcelona, Spain
e-mail: henrik.van.thriel@upc.edu

Key words: 3V's, Applications, Ownership, Security, Privacy

Abstract. This document gives an introduction to what Big Data is, and how Big Data affects us in the normal life. It will describe how Big Data is used in almost every sector of human activity, and describe the discussion concerning ownership when it comes to Big Data and how that affects problems considering security and privacy.

1 INTRODUCTION

We live in an information age. Information is valuable, and data is information. With the possibility of storing almost anything we could use the data to predict the behaviour and interest of persons, but the question will always be for what cost. Does the normal guy on the street actually understand how Big Data affects him? Is it possible that we are generating too much data even for Big Data?

2 INTRO TO BIG DATA

Historically data was generated by workers. Employees and companies were entering data manually into the computers. That changed when the internet started growing and websites like Myspace and Facebook started popping up. Suddenly it was the users that generated the data by writing the data their self. All ready at this moment of time the amount of generated data was growing exponential. Today we have a third level of progression, because now machines are generating data in a way larger scale than the workers and user were able to do. Almost everything on earth is generating data. In fact, according to SINTEF in 2013,ninety percent of the data in the world at that time was created in the two years before [1].

This changed the way we stored the generated data. Before we stored the data in a way that all the data were sent to one single machine/CPU. That means if we want to store more data we must increase the size of the storage. Over time this will could cost an enormous amount of money, and in the end, it will still not be enough storage place. Thats where Big Data makes its entrance. Big data does it the other way around. Instead of sending the data to the processor we take the processors to the data. In that way we could bring an infinite number of cheap processors that we could split the data on. This gives us the possibility to store an endless amount of data without having to pay a fortune.

Most people think about Big Data as a way to store a lot of data, which makes sense given that the most common definition on Big Data is when the amount of data is too much for one machine to process and store. But its the processing of the generated amount of data that really makes Big Data interesting and draws the parallels to artificial intelligence and automation. A general way to explain Big Data is the use of the three Vs; volume, velocity and variety.

- Volume is the quantity of generated and stored data which makes stricter demands on how to process, store and transfer of the data.
- Velocity means that the data is often available in real-time and the analyzing of the data also have to be done in real time.
- Variety describes the difference of data that is being stored. We have to be able store data from different sources like video, audio and images as well as being able to store structured data.

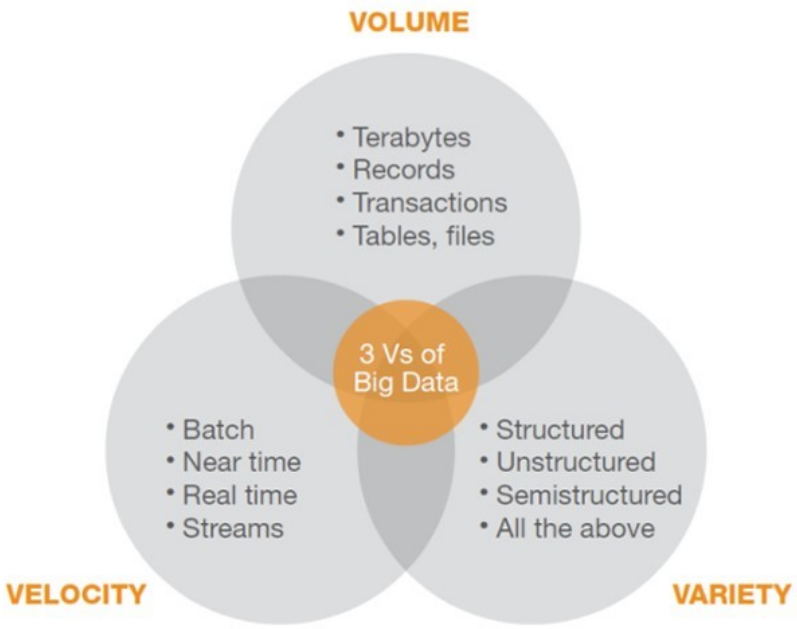

Figure 1: Figure showing the 3V's in Big Data with some characteristics.

3 APPLICATIONS

Big data has a lot of advantages because it offers different properties which makes it to an ideal tool for numerous applications. The benefits are the new kind of data access. It is very easy to collect data in numerous databases and the possibility to use those endless storage of information within seconds. The main benefits are the quick access, the big access, the quick specification and filtering of relevant data, the trustworthiness and the increased security. In addition it is possible to evaluate the data and compare it with others to verify it.

With the data supply it is possible to analyze needed data by combining for example different values to examine a dependence or figure out a special behavior. With the results the making of decision becomes much more easier, because you can quantify everything and look even through complex systems. By making the best decisions the optimization of a process can be very detailed and absolute.

As pointed out before the advantages can be used in almost every sector of human activity.

3.1 Health Sector

Clinical decision support system could support the personnel by Evaluating a lot of data. An example for that is the analyzation of tissue samples to examine the risk of cancer. Right know humans have to check very much indicators by hand and plain sight.

With big data the residues could be examine by analyzing a very high number of samples like pictures in a few seconds. There is also the possibility to do individual analytics applied for a patient profile or to personalize the medicine. The personnel could be paid by their performance, it's possible to analyze disease patterns and improve the public health. Even pandemics could be forecasted.

3.2 Public Sector

In the public sector it is possible to improve the transparency by accessible related data, discover the needs of the people and improve the performance of any actions. The possibility of personalization is a big advantage and simplifies life in general. Big data could personalize the products and services and could take over decision making with automated systems to increase risk. There would be much more innovations with new products and services.

The retail can use the Big data for the analysis of store behavior as well as for the analysis of the variety and prizes optimization's. The product placement design will benefit and the distribution and logistics optimization. For the manufacturing it pens the possibility of demand forecasting, planning of the supply chain and web search based applications.

3.3 Automation

Another big issue is the automatizing process in general. It is possible to develop autonomous driving car by giving access to the data of the environment. The learning process of AI depends on the access to big data. it is possible to create a almost omnipotent AI which can learn from different sectors and analyze the results of actions. With the personal location data we can improve the smart routing, the GEO targeted advertising, urban planning and new business models.

3.4 Environment

The described advantages also allows applications for environmental issues. It is possible to integrate the Big Data to Climate models by overlaying satellite images with global trade databases to estimate the for source of pollution or threatened species wit the global trade database. With those connections unexpected dependencies are pointed out and can be eliminated.

4 WHO OWNS THE DATA?

There is no doubt that Big Data could find solutions to problems we havent been able to solve before, but who is it actually that decides which problems we are going to solve? It is the companies with big resources and the governments who stores and analyze the data. That means again that is not us that decides which problems we are going to solve. We just have to trust that the governments and the companies actually use the concept

of Big Data in the correct way to exploit the potential.

The other important, and difficult question, is who actually owns the data? This is the question that brings up most discuss when talking about Big Data. Every day Twitter sells more than 500 million tweets [2] to companies that tries to analyze the data in order to make money. Are they allowed to sell this information only generated by the users? According to them self, yes. According to others, no.

The ownership discussion brings up another question concerning the security around the storing the data. Big Data has grown so fast that the current laws are outdated when it comes to collecting and analyzing data. Take U.S for example. They dont have one single law that regulate collection, use and sharing of personal information, but different federal and laws. That makes it almost impossible to agree on whats allowed and whats not with the use of Big Data. At the same time when we dont longer control the data we have generated our self we have to trust the ones storing it that it is safe against cyber criminals, also called hackers. In the resent couple of months many companies have been attacked by hackers. And they are not only interested in credit detail, they want every bit of information they could get.

We also have the discussion about what personal information. Everything we do is being tracked, and we have lost the feeling of privacy. The balancing of a persons privacy and the use of Big Data in order to get the most out of the data is one of the most important and difficult aspect of evolution of Big Data. But then again, do we actually care if we are being tracked? Most people push the agree button without reading the conditions to download the app or program. Some people are okay with having less privacy, but it also exist a lot of people that dont know they are losing privacy for every time they push the agree-button.

5 CONCLUSIONS

Connecting data gives knowledge, which could be used to solve problems that we havent been able to solve before. The analyzing of information, the decision making and the optimization is not a new concept. It has existed before, but now we are able to do better analyses, make better decisions and in that way make more complex optimization based on the enormous amount of data that is generated every second.With the new technology we are able to store all the generated in a cost effective way, and with the development of complex analysis we are able to connect everything.

On the contrary, the meaning of Big data isnt just collecting and analyzing as much data as possible. We have to specialize the databases, adapt the policies and strengthen the understanding of how this affects us. There is no other way to deal with the data because right now its not us that store and manage the data, its the companies and the government who has the data, and thereby has the power of using it. Why should they give it away when it gives that amount of power and resources? The only thing we could is to make sure that the power and resources is used for a good purpose.

REFERENCES

[1] SINTEF article
https://www.sciencedaily.com/releases/2013/05/130522085217.htmg

[2] Article about how Twitter earns money
http://www.bbc.com/news/business-24397472