

# The value of big data

by  **Matúš Szepeši,**  
SEO / Data Specialist  
Promiseo s.r.o.

 PROMISEO

**Over 25 years have passed since the Internet was born. In that time, it has become an inseparable part of everyday life. Extremely quick technological development has enabled us to get from an almost-forgotten dial-up connection to the privilege of streaming music and films in 4K, joining a videoconference from home, spending time on social media and playing our favorite games with friends.**

However, technological development has brought not only an improvement of the Internet connection, but also the availability of modern mobile equipment and smartphones: 3,5 billion users use smartphones and over 50% of global traffic comes from smartphones. If we add up data transfers from computers, data centers and everything else connected to the Internet, we end up with an enormous amount of data generated every second.

To illustrate, here are some statistics about how much data are generated on the Internet every minute (2019):

- 2.1 million Snapchats
- 3.8 million Google searches
- 1 million Facebook sign-ins
- 4.5 million videos watched on Youtube
- 188 million emails

Such an amount of information is difficult to process by usual methods. That is why the term Big Data was introduced. The biggest pioneer in the field of Big Data is Google. When it launched, the Internet was relatively small and search engines were not wide-spread. However, the Internet's rapid growth brought along the need for new methods for Google to process a large amount of requests.

## What is Big Data?

Big Data means large volumes of data which need to be processed with a big computing power. To be classified as Big Data, it needs to meet the following conditions:

- Volume – Numerical expression of generated and stored data, usually of values higher than terabytes and petabytes.

- Variety – Data comes in various forms – structured (relational databases, spreadsheet files, etc.), non-structured (with no clear format – satellite images, messages, videos), and semi-structured (they seem non-structured at first sight, however, they have features which make them easier to process).
- Velocity – The speed at which the data are generated and processed.
- Veracity – The truthfulness or reliability of the data, which refers to the data quality and data value.

Since the amount of Big Data increases exponentially – more than 500 TB of data are uploaded to Facebook Database alone, in a single day – it represents a real problem in terms of analysis.

## Why should we care about Big Data?

Now, you may be thinking that if Big Data is so problematic, why is everyone so obsessed with it? The answer lies in the benefits it offers.

First of all, it is the value Big Data can provide by its processing and analysis. For example, Netflix collects data about more than 100 million of their customers. These help the network to understand what each customer wants to watch. As a result, the customer is satisfied because they get what they want without having to look for it. Netflix is satisfied because it keeps its customers. Similarly, credit card companies collect real-time data about when and where the cards were used. Imagine that a credit card was used at point A. Three hours later, the same card was

used at point B which is 5,000 km from point A. It is not possible for a person to travel such a distance in such a short time. Therefore, the credit card company is able to find out that somebody is trying to abuse their system.

Big Data opens possibilities to companies in various fields, i.e. banking, medicine, media, advertisement, transportation, manufacture, wholesale and retail.

## Cloud computing and Big Data

Your company might also have a large amount of data which has the potential of bringing you larger profits. How to process it? The answer lies in parallel data processing, i.e. cloud computing. Depending on the amount of data, thousands of computers might work on its processing while each computer is focused only on a certain amount of data. This process is fundamentally faster than having data processed by a single computer.

## Why is cloud computing a better choice?

To put it simply, you have two choices: either to create a local system (on-premises) for data processing, or use the cloud services. If you are not planning to rent a local system in the future, there is not much point in building it. Here are the main reasons why it is more efficient to use a provider rather than creating an on-premises system:

### Scalability:

- Building and widening your own solutions is costly and your possibilities are limited. Once you do not need extended computing abilities, it is difficult to reduce, which

brings financial losses.

- Cloud computing – you only pay for the time used in the service. Moreover, it is easier to add or reduce the computing ability.

### Server Storage:

- On-premises – with your own solution, you need to physically build your server, take care of its operation and cover the energy costs.
- Cloud computing is offered by cloud service providers who take care of the servers and save you money and space.

### Data security a data loss

- Taking care of security and sustainability of data is difficult and requires further costs.
- Cloud services have encrypted servers and data are stored on more discs in case one of them is damaged.

Google, Microsoft and Amazon are amongst the most famous companies offering public clouds available to the general public. There is also a possibility of using a private cloud which is managed by a single company with strict security measures.

## Conclusion

If you own a company, there will come a time when you need to process and analyze your company's data. You may start to look for possibilities of using cloud computing and process the gigabytes of your data with the aim of increasing your profits or being one step ahead of your competitors.

Smaller companies which do not have large amounts of data might analyze their "small data" via spreadsheets and Business Intelligence tools. Marketing agencies offer the service of company reporting which includes processing and analyzing data for small and medium-sized companies with the aim of clarifying internal processes, improving management and using company data to increase its competitiveness, allow better and faster decision-making and secure higher profits.