

BIG DATA

Tiunchik P. D.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Ukrainets N. G. – Senior Lecturer

Annotation. The article deals with the term of Big Data. It will cover Big Data's definition and characteristic. It will also cover the benefits which can be gained by integrating big data analysis into various spheres of life. This article will also touch the subject of data privacy. This issue is relevant in this day and age because big data analysis can turn useless chunks of data into valuable information and, eventually, profit.

Keywords. Big data, volume, velocity, variety, veracity, value.

In the modern world, where more than a half of the Earth's population are active Internet users, a great deal of data is gathered and processed every day. You'd be surprised at how much data an average mobile phone user generates daily, in the form of texts, photos, videos, searches, phone calls, music, and many more. All this material is gathered by private companies, governments, and many other establishments for the purpose of structuring, managing, processing, and extracting useful information and statistics from it. All those processes are done by qualified professionals from the field of big data.

What is 'big data'? This term can have two distinct definitions:

Big data is a field that treats ways to analyze, systematically extract information from, or otherwise, deal with data sets that are too large or complex to be dealt with by traditional data-processing application software [1].

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is data with so large size and complexity that none of the traditional data management tools can store it or process it efficiently [3].

So, both the material and the field itself can be described by this term.

And now some words about the history of Big Data. The term has been in use since the 1990s, after John R. Mashley, a U.S. computer scientist, director, and entrepreneur popularized it [2].

In order to more easily tell big data and usual data apart, several characteristics are used to define big data

Volume – describes the size/quantity of generated and processed data. The size of big data is usually greater than petabytes (1024000 gigabytes). The volume of big data makes it impossible to process it using traditional methods and technologies. In other words, a simple laptop or a desktop processor isn't able to deal with big data sets. Daily credit card transactions can be given as an example of a high-volume data set. The most valuable data sets are always huge in size, as it means that it is most exhaustive and accurate.

Velocity – the speed at which data is generated and processed. Big data is usually produced continually and available in real-time. Obviously, outdated data is almost useless, so data sets should always be gathered at a fast pace. Social Media posts are an example of high-velocity data.

Variety – the type and nature of the data. Big data comes from a variety of sources and can be one of the three types: structured (having an expected set of file formats and types, the easiest type of big data to process), unstructured (not having an expected set of formats, harder to process) and semi-structured (the combination of the two). During earlier days, spreadsheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications [3]. CCTV audio and video files from various parts of a city are an example of a high-variety data set.

Veracity – the accuracy and quality of analyzed data. High-veracity data is valuable to analyze and has records that contribute to the overall results. Low-veracity data has a lot of useless data, nicknamed "noise" in the data processing field. Data from a medical trial or examination is an example of high-veracity data.

Value – the benefits or profit, which can be gained from working with data. This characteristic makes it worthwhile for establishments to study the methods and intricacies of working with big data. Processing big data always yields good results.

In short, big data has to be great in size, generated and processed quickly, gathered in many forms from many sources, low in useless data, and yield valuable results to be truly called "Big Data".

These are only the five main characteristics used to describe big data, and there are several other possible characteristics. Originally, there were only three characteristics, and with time their number grew to accommodate for the growing quantity of data gathered and the evolution of data-processing software and applications. These characteristics are sometimes referred to as "the three/four/five V's".

Big data processing has its use in many spheres of life, ranging from healthcare and crime prediction to advertisement and customer service. To give a clearer picture of its application, let's take social networking websites like Facebook and VK. Facebook has a user base of 3 billion monthly active users, which means that the volume and the velocity of gathered and produced data are great. The variety of material acquired from social networking websites come from searches, visited communities, private messages, viewed media, and many other activities of every user. Such data is relatively trustworthy, as it can tell the user's interests, needs and wishes, it has veracity. To Facebook, the value of processed big data is undoubtedly great as it lets them show targeted advertisements to users, recommend new communities to join, implement changes and updates to the functionality of their website and just have better customer support in general.

With a smart and effective way for companies to gather and benefit from user data, some users worry how this change in information management can negatively impact their lives. There have already been cases when a person's social media activity directly impacted their chances of getting a loan from a bank [4]. Today, it's not even required to sign shady deals with shady companies to get your hands on a person's crucial information. In this day and age, it's enough to just visit a person's social media account to learn a great deal about them: their age, place of living, friends, family, thoughts, job stability, and many others. Private companies, which specialize in dealing with big data acquired from open sources get requests from banks, law enforcement, entrepreneurs, etc. to gather and analyze information about borrowers, potential criminals, or customer bases. Many important and beneficial decisions can be made just from the person's public data that is available to everyone, now imagine if social media companies themselves decide to start selling your private data to gain profit. Fortunately, government authorities usually try to keep an eye out for such cases, like when Mark Zuckerberg, chairman of Facebook, was interrogated by lawmakers, on suspicion of manipulating information and selling the users' private data [5].

With the development of hardware and software, the ways of dealing with big data will be only improved further, which will greatly affect many aspects of our daily lives. Despite rising concerns about Internet privacy, it is evident that the phenomenon of big data is here to stay and develop into a more profitable field, and the degrees of 'Data Analyst', 'Data Scientist' and 'Data Specialist' will be sought after.

References:

1. Big Data [Electronic resource]. – Mode of access: https://en.wikipedia.org/wiki/Big_data. – Date of access: 14.03.2021.

57-я научная конференция аспирантов, магистрантов и студентов БГУИР, 2021 г

2. John Mashley [Electronic resource]. – Mode of access: https://en.wikipedia.org/wiki/John_Mashey. – Date of access: 14.03.2021.

3. What is BIG DATA? Introduction, Types, Characteristics, Examples [Electronic resource]. – Mode of access: <https://www.guru99.com/what-is-big-data.html#3>. – Date of access: 14.03.2021.

4. Lenders use social media to screen borrowers - WSJ [Electronic resource]. – Mode of access: <https://www.wsj.com/articles/borrowers-hit-socialmedia-hurdles-1389224469>. – Date of access: 15.03.2021.

5. 2 days, 100 hours, 600 questions: What happened when Mark Zuckerberg went to Washington [Electronic resource]. – Mode of access: <https://www.nytimes.com/2018/04/12/technology/mark-zuckerberg-testimony.html>. – Date of access: 15.03.2021.