DEVELOPMENT ON ADVANCED TECHNOLOGIES – DESIGN AND DEVELOPMENT OF CLOUD COMPUTING MODEL

Alexandra Briasouli
PhD Candidate[1]

Daniela Minkovska
PhD, Associate Professor[1]
daniela@tu-sofia.bg

Lyudmila Stoyanova
PhD, Associate Professor[1]
lstoyanova@tu-sofia.bg

Department of Computer Systems and Technologies
Technical University of Sofia
8 Kliment Ohridski str., Sofia, Bulgaria, 1000

Abstract: Big Data has been created from virtually everything around us at all times. Every digital media interaction generates data, from computer browsing and online retail to iTunes shopping and Facebook likes. This data is captured from multiple sources, with terrifying speed, volume and variety. But in order to extract substantial value from them, one must possess the optimal processing power, the appropriate analysis tools and, of course, the corresponding skills. The range of data collected by businesses today is almost unreal. According to IBM, more than 2.5 times four million data bytes generated per year, while the amount of data generated increases at such an astonishing rate that 90 % of it has been generated in just the last two years. Big Data have recently attracted substantial interest from both academics and practitioners. Big Data Analytics (BDA) is increasingly becoming a trending practice that many organizations are adopting with the purpose of constructing valuable information from BD. The analytics process, including the deployment and use of BDA tools, is seen by organizations as a tool to improve operational efficiency though it has strategic potential, drive new revenue streams and gain competitive advantages over business rivals. However, there are different types of analytic applications to consider. This paper presents a view of the BD challenges and methods to help to understand the significance of using the Big Data Technologies. This article based on a bibliographic review, on texts published in scientific journals, on relevant research dealing with the big data that have exploded in recent years, as they are increasingly linked to technology.

Keywords: Data, big data, knowledge mining, information explosion, data management, social networks.

1. Introduction

In recent years there has been a growing interest in Big Data Management (or Big Data) driven by real production needs. The term first appeared in 1997 by NASA scientists. They reported that they were unable to graphically visualize the data sets they owned, as they were so large that it was impossible to store them in main memory, on the local disk, and on an external hard disk. So they said they were having a Big Data problem. The latest technological developments, mainly in the field of communications and integrated circuits, have made it possible to create mechanisms for monitoring the operations of an organization at a very detailed level. This detailed digitization of production processes has made large organizations as well as small companies capable of producing huge volumes of data at a very fast pace [1].

Intel is one of the few companies to provide quantitative data in their literature. Intel associates big data with organizations that «generate an average of 300 terabytes (TB) of data per week» [2, 3]. He claims that the most common type of data encountered is business transactions stored in relational databases (as defined by Oracle), followed by documents, e-mails, blogs and social media.

Finally, it would be wrong to make a precise definition of the term Big Data, since every company uses the science of Big Data for different purposes, so their multidimensional nature is obvious, which is clearly constantly enriched over time.

2. Methods

To capture the value from Big Data, it is necessary to develop new methods and technologies for analyzing it. Until now, scientists have developed a wide variety of techniques and technologies to capture, curate, analyze and visualize Big Data. Even so, they are far away from meeting variety of needs. These techniques and technologies cross a number of discipline, including computer science, economics, mathematics, statistics and other expertise. Multidisciplinary methods are needed to discovery the valuable information from Big Data. Let's discuss current techniques and technologies for exploiting data intensive applications. We need tools (platforms) to make sense of Big Data. Current tools concentrate on three classes, namely, batch processing tools, stream processing tools, and interactive analysis tools.

Big Data methods involve a number of disciplines, including statistics, data mining, machine learning, neural networks, social network analysis, signal processing, pattern recognition, optimization methods and visualization approaches. There are many specific techniques in these disciplines, and they overlap with each other hourly.

Optimization Methods have been applied to solve quantitative problems in many fields, such as physics, biology, engineering, and economics. In [4], several computational strategies for addressing global optimization problems are discussed, such as simulated annealing, adaptive simulated annealing, quantum annealing, as well as genetic algorithm, which naturally lends itself to parallelism and therefore can be highly efficient. Stochastic optimization, including genetic programming, evolutionary programming, and particle swarm optimization are useful and specific optimization techniques inspired by the process of nature. However, they often have high complexity in memory and time consumption. Many research works [5–7] have been done to scale up the large-scale optimization by cooperative co-evolutionary algorithms. Real-time optimization is also required in many Big Data application, such as WSNs and ITSs. Data reduction [8] and parallelization [9] are also alternative approaches in optimization problems.

3. Results

The «big data» marks the moment when the «information society» finally comes to fulfill the promise that its name implies. Data acquires a leading role. All these digital pieces that have been assembled can now be managed in innovative ways to serve new purposes and add value to things (**Fig. 1**) [10].

«Big data» allows to experiment faster and discover new ways. These advantages will lead to more innovations. But sometimes, the spark of discovery sparks from the silence of the data. This is something that no amount of data can ever confirm or document, since they do not yet exist [11–13].

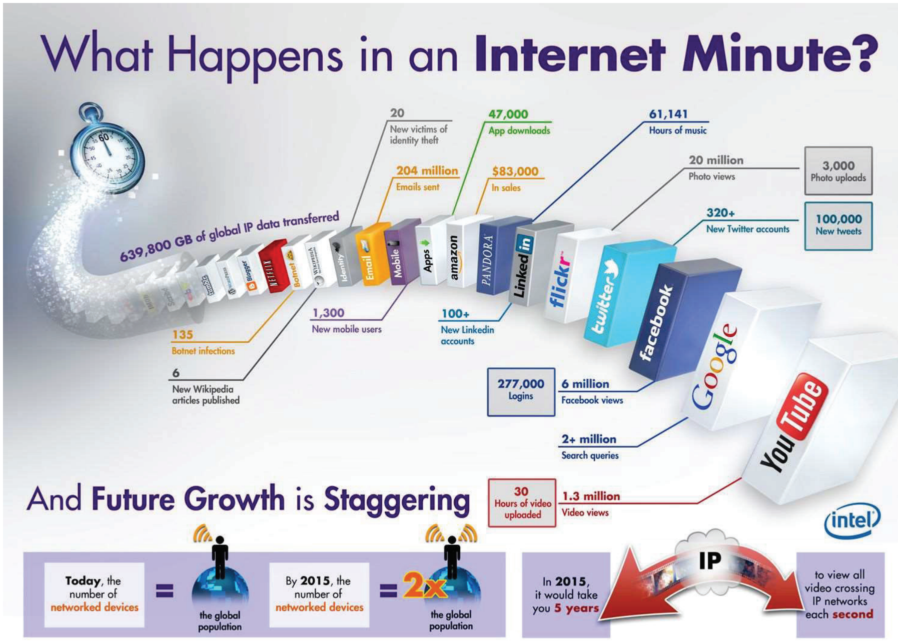

Fig. 1. Volume of data generated per minute

4. Discussion

«Big data» is a source and a tool. They are meant to inform rather than explain. They point in the direction of an explanation, but they can also lead to misunderstanding: it depends on how well you process them. And as amazing as the power of «big data» may seem, its seductive brilliance should never blind us to its inherent imperfections. It is better to adopt this technology, accepting not only its power but also its limits. The ability to capture personal data is often contained within the very tools let's use every day, from websites to mobile applications.

Given the myriad of ways in which information can be reused, rearranged and sold, it is often impossible for users to consent or refuse «innovative uses» of data that were unimaginable when they first began to collect this data [14, 15].

Big data hides valuable knowledge as their analysis can lead to significant optimizations of production but also problems, since the existing technological solutions for data management do not fully meet the volume and their nature. Digital data is now found everywhere: in every sector, in every economy, in every organization and user of digital technology. Big data is increasingly gaining the interest of leaders from all sectors, while consumers of products and services are expected to benefit from their exploitation. The ability to store, collect, combine data and use the results to perform detailed analyzes has become much more accessible and feasible. For less than $ 600, one can buy a drive capable of storing all of the world's music. Data mining tools are also significantly improved, as the software available to implement increasingly complexity techniques is combined with increasing computing power. In addition, the ability to generate, communicate, share and access data has been boosted by the growing number of people, devices and sensors currently connected to digital networks. In 2010, more than 4 billion people, or 60 percent of the world's population, used cell phones, and about 12 percent of those people had smartphones, whose penetration is growing by more than 20 percent a year. More than 30 million networked sensor nodes are now located in the transport, automotive, utility, and retail sectors. The number of these sensors is increasing by more than 30. Many technological innovations have led to a dramatic increase in data and data collection. This is why large-scale data has become a recent area of strategic investment for IT organizations [16].

In 2012, the Obama administration announced the Big Data Research and Development Initiative, which consists of 84 programs in six areas. The report predicted that by 2020 the digital world would hold 40 zettabytes, 57 times the total number of grains of sand from all the world's beaches (**Fig. 2**).

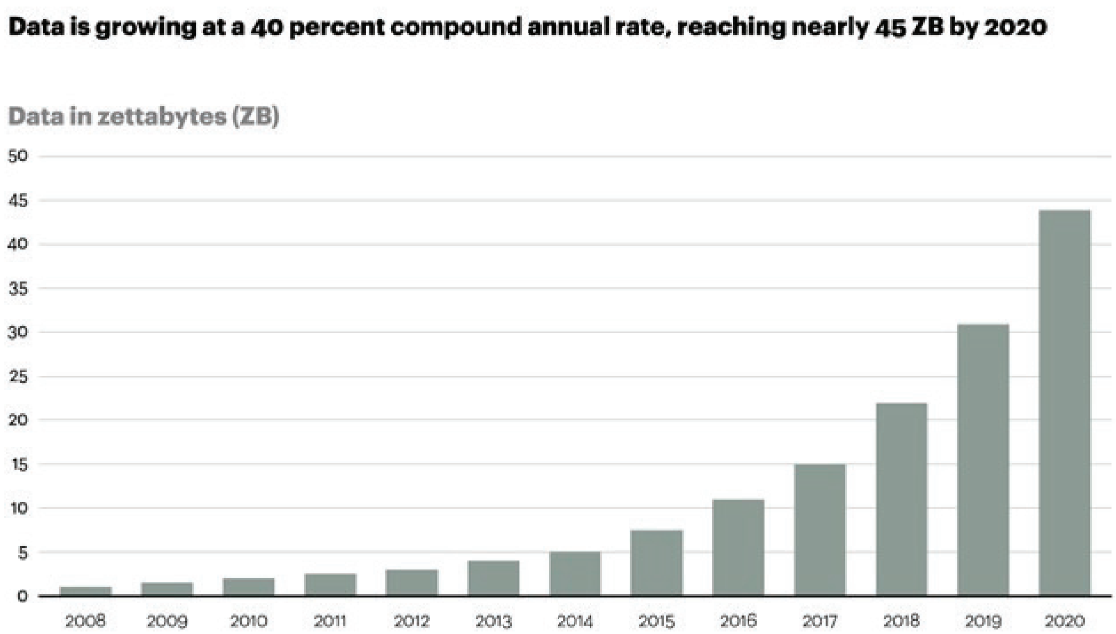

Fig. 2. Percentage data increase – Forecast until 2020

5. Conclusions

But it is not just organizations that produce huge volumes of data. Even on a smaller scale of organization, at the individual level, data production is unprecedented [17, 18]. Most people have a digital self as a display of their activities on social media. Google estimates that every two days the digital material created by users is the same size as the printed material produced by mankind from the beginning of writing until 2003 [19]. An explosion in the volume of data produced is still observed in scientific research. Fields such as medicine, astronomy, meteorology and biology, thanks to new technologies, new telescopes, new and inexpensive sensors and new DNA expression machines, can and do produce volumes of data that cannot be existing infrastructure [20, 21].

In fact, we observe the growth rates that are exponentially distributed. We would say, however, that there is no data size limit above which they are called «Big Data». It is estimated that today with this term we usually refer to volumes of data ranging from a few terabytes to tens or even hundreds of petabytes (1,024 terabytes) or exabytes (1,024 petabytes) or zetabytes (1,024 exabytes). Thus, an even bigger «information explosion» is predicted for the coming years.

References

1. Sivinski, G., Okuliar, A., Kjolbye, L. (2017). Is big data a big deal? A competition law approach to big data. European Competition Journal, 13 (2-3), 199–227. doi: https://doi.org/10.1080/17441056.2017.1362866
2. Matturdi, B., Zhou, X., Li, S., Lin, F. (2014). Big Data security and privacy: A review. China Communications, 11 (14), 135–145. doi: https://doi.org/10.1109/cc.2014.7085614
3. Smith, M., Szongott, C., Henne, B., von Voigt, G. (2012). Big data privacy issues in public social media. 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST). doi: https://doi.org/10.1109/dest.2012.6227909
4. Sahimi, M., Hamzehpour, H. (2010). Efficient Computational Strategies for Solving Global Optimization Problems. Computing in Science & Engineering, 12 (4), 74–83. doi: https://doi.org/10.1109/mcse.2010.85
5. Li, X., Yao, X. (2012). Cooperatively Coevolving Particle Swarms for Large Scale Optimization. IEEE Transactions on Evolutionary Computation, 16 (2), 210–224. doi: https://doi.org/10.1109/tevc.2011.2112662
6. Del Valle, Y., Venayagamoorthy, G. K., Mohagheghi, S., Hernandez, J.-C., Harley, R. G. (2008). Particle Swarm Optimization: Basic Concepts, Variants and Applications in Power Systems. IEEE Transactions on Evolutionary Computation, 12 (2), 171–195. doi: https://doi.org/10.1109/tevc.2007.896686
7. Yang, Z., Tang, K., Yao, X. (2008). Large scale evolutionary optimization using cooperative coevolution. Information Sciences, 178 (15), 2985–2999. doi: https://doi.org/10.1016/j.ins.2008.02.017
8. Yan, J., Liu, N., Yan, S., Yang, Q., Fan, W., Wei, W., Chen, Z. (2011). Trace-Oriented Feature Analysis for Large-Scale Text Data Dimension Reduction. IEEE Transactions on Knowledge and Data Engineering, 23 (7), 1103–1117. doi: https://doi.org/10.1109/tkde.2010.34
9. Yao, W., Chen, X., Zhao, Y., van Tooren, M. (2012). Concurrent Subspace Width Optimization Method for RBF Neural Network Modeling. IEEE Transactions on Neural Networks and Learning Systems, 23 (2), 247–259. doi: https://doi.org/10.1109/tnnls.2011.2178560
10. Philip Chen, C. L., Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314–347. doi: https://doi.org/10.1016/j.ins.2014.01.015
11. Chen, M., Mao, S., Liu, Y. (2014). Big Data: A Survey. Mobile Networks and Applications, 19 (2), 171–209. doi: https://doi.org/10.1007/s11036-013-0489-0
12. He, Y., Yu, F. R., Zhao, N., Yin, H., Yao, H., Qiu, R. C. (2016). Big Data Analytics in Mobile Cellular Networks. IEEE Access, 4, 1985–1996. doi: https://doi.org/10.1109/access.2016.2540520
13. Kim, G.-H., Trimi, S., Chung, J.-H. (2014). Big-data applications in the government sector. Communications of the ACM, 57 (3), 78–85. doi: https://doi.org/10.1145/2500873
14. Hu, H., Wen, Y., Chua, T.-S., Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access, 2, 652–687. doi: https://doi.org/10.1109/access.2014.2332453
15. Kitchin, R., Lauriault, T. P. (2014). Small data in the era of big data. GeoJournal, 80 (4), 463–475. doi: https://doi.org/10.1007/s10708-014-9601-7
16. Goudarzi, M. (2019). Heterogeneous Architectures for Big Data Batch Processing in MapReduce Paradigm. IEEE Transactions on Big Data, 5 (1), 18–33. doi: https://doi.org/10.1109/tbdata.2017.2736557
17. Bertino, E. (2015). Big Data - Security and Privacy. 2015 IEEE International Congress on Big Data. doi: https://doi.org/10.1109/bigdatacongress.2015.126
18. Strang, K. D., Sun, Z. (2017). Big Data Paradigm: What is the Status of Privacy and Security? Annals of Data Science, 4 (1), 1–17. doi: https://doi.org/10.1007/s40745-016-0096-6
19. Dolev, S., Florissi, P., Gudes, E., Sharma, S., Singer, I. (2019). A Survey on Geographically Distributed Big-Data Processing Using MapReduce. IEEE Transactions on Big Data, 5 (1), 60–80. doi: https://doi.org/10.1109/tbdata.2017.2723473

20. Costa, F. F. (2014). Big data in biomedicine. Drug Discovery Today, 19 (4), 433–440. doi: https://doi.org/10.1016/j.drudis.2013.10.012

21. Fan, J., Han, F., Liu, H. (2014). Challenges of Big Data analysis. National Science Review, 1 (2), 293–314. doi: https://doi.org/10.1093/nsr/nwt032

Received date 14.10.2021
Accepted date 15.11.2021
Published date 29.11.2021

How to cite: *Briasouli, A., Minkovska, D., Stoyanova, L. (2021). Development on advanced technologies – design and development of cloud computing model. Technology transfer: fundamental principles and innovative technical solutions, 13–16. doi: https://doi.org/10.21303/2585-6847.2021.002228*