https://doi.org/10.5281/zenodo.4289387 **Review Article**

THE ROLE OF DATA SCIENCE IN CHANGING THE BUSINESS LANDSCAPE

Computer Science

Keywords: data science, data analysis, business, data management.

Mefail Tahiri	**University of Tetova "Fadil Sulejmani". North Macedonia.**
Ejup Rustemi	**University of Tetova "Fadil Sulejmani". North Macedonia.**

Abstract

Businesses come in different scale and scope. Almost every day we hear about a new company that has emerged somewhere in the world. Sometimes we learn that a friend of ours has decided to open his or her own business, or at times, ourselves might be tempted to undertake such a task. Whatever the case, one thing is for certain, we will have to deal with data which will have to be analyzed so we can understand how customers behave, what products people love or hate; this is the point when Data Science comes in help, by providing tools so we can extract the needed information to properly develop our business. This is of high importance if we can understand that we are living in a time when change is unavoidable and requires us to be always prepared. In this paper, we will try to evaluate this aspect in believing to provide a general overview regarding this very important issue.

Introduction

Byte, kilobyte, megabyte, gigabyte, terabyte, petabyte, and so on; each of these terms suggests storage space where we can keep our data to retrieve them, analyze them, and act upon them based on our personal needs or the needs of our business. A large amount of data automatically means complex analytics, but a large amount of data doesn't have to mean something that cannot be managed properly. Data Science is the key to all of this complexity, where data scientists use new technologies to visualize and contextualize any given data set. As each day goes by, data becomes bigger and bigger and we come to the point where storage becomes an issue, especially when we are dependent on our own data storage. To avoid these problems the majority of enterprises are using cloud storage and data management tools, something which is provided by the largest IT companies in the world such as Microsoft, Oracle, Google, and Amazon. Data Science is the nexus around which data management is being executed in an ever-changing business landscape.

The Data Scientist

If we were to define what a data scientist is, we would declare that he or sheexpertizes in data analytics. If we leave it at this, we are downgrading the importance and scope of what it means to be a data scientist. One such expert knows how to use tools and skills from both information technology and social sciences. A data scientist, by analyzing the data, he encounters human behavior, so he must be aware of how humans as social beings, react in certain market circumstances. So, contextualizing things is the key to understanding data. Data scientists must use any statistical data from previous marketing, have an open mind towards them, at times he

may be skeptical of any existing assumptions, but always prepared to find proper solutions to any probable business challenge.

Such an expert must be aware of security issues that may affect the business. A data breach is one of the most important issues. Attack on data doesn't always have to come from individual cyber attackers, they can also come from other competing companies, who always want to be ahead of you, and want to gain knowledge about your future plans. As we all know, future plans are made upon knowledge gained from data analytics, and breaching them means trouble for your business from a profitable point of view. One other thing that we must notice is that there also reputational issues that need to be taken seriously. A lot of times, companies keep personal data from their customers, and sensitively using those data is very important, because if they are breached or if we use them for a purpose other than for what we actually need, will certainly hurt our reputation. The following figure gives a visual representation of what a data scientist should be aware of.

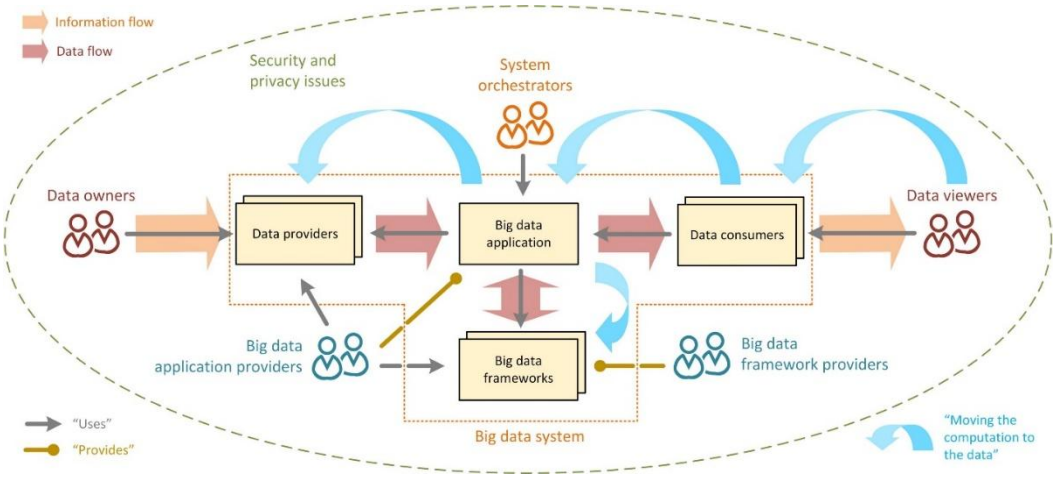

Fig. 1 *What a Data Scientist should know*[1]

Why Data Science?

Some statistics show that by this year (2020) data administrators will have to handle around 40 zettabytes of data. If that doesn't seem big enough, as a comparison, it is more than 300 times than what they had to deal with in 2005. So what does it mean? First and foremost it shows that a huge amount of data still isn't converted into business knowledge upon which people can act. Secondly, the enormous increase in data is obvious that it has to do with the Internet. People produce two and a half quintillion bytes of data every single day. And thirdly, a data scientist should find ways and tools to manage this data accordingly (Vermeulen, 2018).

Data Science is a step by step, level by level, approach to data management. The following picture displays each level of data processing.

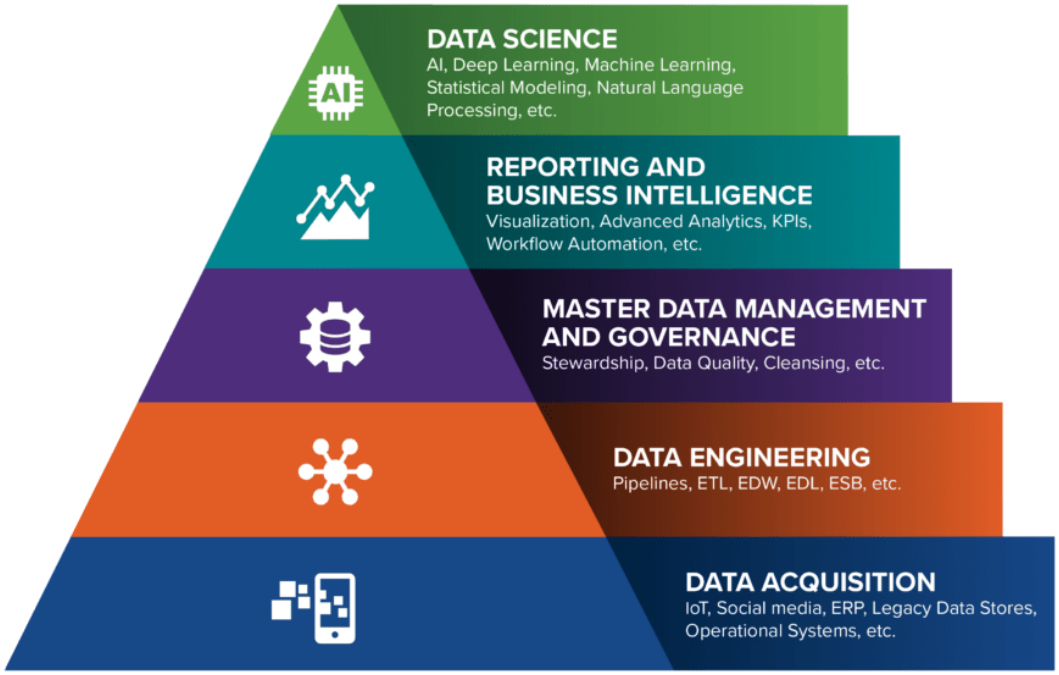

Fig. 2 *Data Science Pyramid*[2]

From the above pyramid, we can see that Data Science is at the highest level of data processing because it provides a different kind of approach to data statistics and analytics.

From the picture, we can see that the lowest level represents the sources from where data is acquired. It can be any kind of source ranging from ERP, Social Networks, small or large businesses, etc.

The next level has to do with how we engineer the acquired data, which is done by Pipelines, ETL technique (Extract, Transform, Load), Data Warehousing, etc..

The third level has to do with checking data quality and cleaning any unimportant data which although insignificant, it can still take valuable time to analyze.

The fourth level is specifically important because it deals with reporting and business intelligence. This is the step we start visualizing data so we can have a better understanding of it, we begin more advanced analytics and workflow automation that provides automated actions in the business process.

The final level is Data Science, that provides the most recent methods and techniques such as Artificial Intelligence, Deep Learning, etc..

Conclusion

Data should be managed, that is something everyone agrees, but when it comes to the tools, we encounter differences in what is being used based on the needs of a given company. We have explained that Data Science is the future of data management, still, we must be aware that not every enterprise can afford that. It's not just a matter of finances, but it is more a problem of lacking the proper expertise, of people who are prepared to undertake such tasks that belong to Data Science. Things are going to change in the future, at least we hope so, and experts will certainly be in a greater number, but we should be aware that the business landscape is changing fast and if we want to be competitive in such an environment our actions must be fast and appropriate.

References

Godsey, B. (2017); Think Like a Data Scientist: Tackle the data science process step-by-step. Manning Publications, Shelter Island.

Patil, DJ. (2011); *Building Data Science Teams*. O'Reilly Media, Sebastopol.

Peng, R.D., Matsui, E. (2015); *The Art of Data Science*. Leanpub., British Columbia.

Skiena, S.S. (2017); *The Data Science Design Manual*. Springer, New York.

Squire, M. (2015); *Clean Data - Save time by discovering effortless strategies for cleaning, organizing, and manipulating your data*. Packt Publishing, Birmingham.

Vermeulen, A.F. (2018); *Practical Data Science: A Guide to Building the Technology Stack for Turning Data Lakes into Business Assets*. Apress, West Kilbride North Ayrshire.

Weblinks

[1] https://sensecorp.com/data-science-pyramid/

[2] https://www.kdnuggets.com/2019/10/data-scientist-data-management.html