

THE IMPACT OF DATA SCIENCE ON DATA MANAGEMENT

Doc. Dr. Ejup Rustemi¹

Prof. Dr. Mefail Tahiri²

¹University of Tetova, ejup.rustemi@yahoo.com

²University of Tetova, mefailt@gmail.com

Abstract

In our time data expert need to deal with an enormous amount of data that need to be managed and build knowledge upon them to maintain a solid step in the landscape of continuous development and change. The lack of proper data management will certainly lead towards confusion, problematic issues, stress and, of course, failure. If we take all of this into consideration, it is quite clear that managers need to find tools with which they can facilitate such an undertaking so they can achieve the objectives of a given enterprise, no matter its size and scope. Data Science is an IT field that tries to provide how managers can achieve these tasks, something that we will try to exemplify and evaluate in our paper. In order to prepare and evaluate the outcome of the paper we have used various methods that have provided the means to analyze the problem from different aspects. First and foremost, the exploratory method has been used to explore the huge material available in books, websites, other paper of the field, all of which have been contextualized and selected to give the needed clarification of the problem and to make the point as concise as possible. Considering the importance of data and the value it brings to the table of any give business or any other institution makes it obvious to that managing it is the key to success, which is why we have chosen this topic and our purpose is to make people focus on proper data management.

Keywords: data science, management, data, information.

Introduction

Every one of us needs, at some point, to manage some data. Individuals will certainly have a lot less work to do than let's say a medium organization, still, data management is vital to how we do our work. Data science is the bridge upon which organizations are building their managing environment with the tools provided to achieve their goals in any given case and situation. It doesn't matter if it is a small company at your neighborhood or a large-scale corporation such as Microsoft, Google or Apple, each one of them must manage their data to see what works and what doesn't regarding their products, and what may be even more important, it gives them the tools to make strategic moves for the future.

Importance of Data Management

We are continuously witnessing a rapid growth of data no matter their nature or importance. Whatever their purpose maybe they represent a task with which data managers need to deal with so they can obtain a better understanding where their business stands, how it can be improved and where lies the future. Will all this end? Probably not, because data management and analytics are growing daily, databases in general and data management, in particular, will continue to have even larger importance in every field of development.

Data managers will always strive to get meaning out of a huge pile of data so they can use it in their "war" with the competition. In other words, this is the only way a company can survive in an ever-changing landscape. It is not unusual to see many companies use software that will alert them about any probable lack of products or negative customer response for any given product or service. This is of very high importance if a company wants to know how to plan their strategy for the future. This is the most important issue for every company, to know what to expect tomorrow, because a good manager or database administrator always is directed towards the future, towards what can the next day bring and the next tool help in the business landscape.

As we can see, we are dealing here with an issue of utmost importance which is very important but also very sensitive too. The main problem is with finding competent people who can deal with such task which requires specific abilities to see and plan the future. Certainty can never be achieved and that is a fact, but working towards a data analysis-based strategy will certainly create a better environment with as fewer problems as possible. Working in this direction will require specific training and specific tools, which in return will require a certain

amount of expenses. The field of Data Science tries to provide the means to deal with such things in providing a framework upon which every expert can bring to the table his or her expertise, and together they can reach the objectives.

We mention all of this not to demotivate young people who want to get into this field of technology, but on the contrary, to prepare them and make them aware that these professions require a serious approach and specific competencies and abilities. If we take into account the profession of Database Administrator, we will discover that it is one of the most requested jobs in any given country, suggesting that people who are capable of undertaking such a job are not in great number. If this is true for the more advanced countries, particularly in the field of Information Technologies, the Balkans, where we work and learn ourselves the issue becomes even more problematic, considering we are late in the application of such tools. So, data management will require us to train talented people, provide them with the proper tools and build a working environment for them, so they can work and reach the needed results. (Hoffer, Ramesh, Topi; 2011)

Data Science we can be considered a new approach in dealing with IT issues, particularly in the field of data management, displays a new approach to data analysis to better understand data and generate important information to be successful in the world of business, especially compared to the competition.

The complexity of data management derives from the fact that a company, especially if it is new, will be faced with a wide range of applications, some of which will be desktop-based, some web-based, and particularly nowadays, mobile-based, each of which, although it may be similar in content, still, it will require a different approach to solve a given task. Data Science tries to build a framework for a seamless transition from one environment to the other, or building applications that will look and act similar no matter the device from which a database administrator will do his or her job.

Data management requires teamwork. No single person can analyze the data from even a small enterprise. Although applications facilitate a high degree this process, still, there must be human intervention, especially when we reach the phase of reporting. With the advance of Artificial Intelligence, these tasks are becoming more easily managed, but we're still not at a level of full automation (Hoffer, Ramesh, Topi; 2011).

Data management represents a complex network of cooperation, analysis, evaluation, assessment and decision-making. Making a decision is the most important thing in a given company; but good decision derives from good data analytics, therefore database management is crucial to make better decisions.

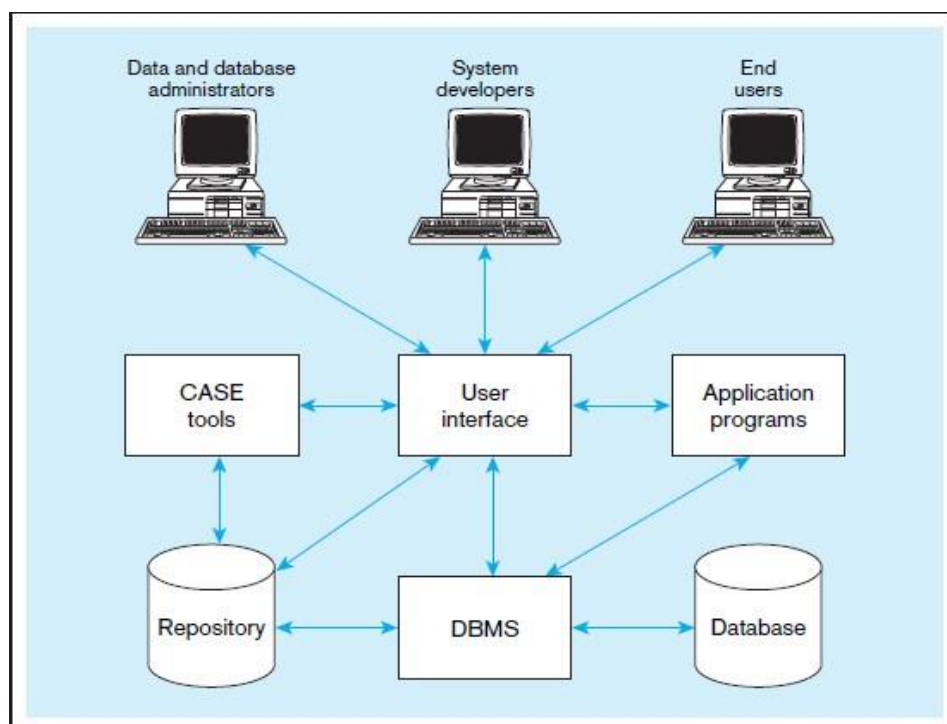


Fig. 1 *Components of the database environment*

Databases come in different kinds and scale. Each one of us has our own databases; it can be inside our laptops, smartphones, desktop PC-s, tablets, etc. We make folders where we keep our personal documents; we save our contacts in our phones, each of these represent data that most of the times our devices arrange and manage by themselves, rarely requiring our intervention. However, when we deal with large corporations, with an enormous amount of data, things become more difficult and require different measures of management and analysis because here we are dealing with enterprise-customer relationship, product sales analysis, services analysis and so on.

When we reach the level of sharing data, we encounter databases known as Two-Tier Client/Server Databases. These databases involve groups, some small, some large, but usually around 25 people, all of who exchange data, provide feedback, enter new data, modify previously entered data. All of this makes data management more prone to problems and needs a specific approach to deal with such tasks (Hoffer, Ramesh, Topi; 2011).

In other words, to manage data, means to acquire new information, acquiring new information helps in making proper decisions.

Gathering experts to teach talented people who want to do such a job as managing data is the first step towards building a good strategy so we can be able enough to face the growing challenges of enterprise development and change because there is no doubt that if we manage everything properly, with skilled personnel, our enterprise will be ready to advance in this world of high competition.

What about Data Analytics

When we analyze data, one of the reasons we do such a task is to evaluate what has been done up to that point. Data analysis will also provide us with the means to evaluate what should be done next. Data Science in providing tools such as Tableau, Microsoft PowerBI, etc., gives data analysts the groundwork upon which the enterprise will build the strategy, usually for a year. All of this depending on the size of a particular enterprise.

A lot of times, some organizations will need to build their strategy from the ground up, without the help of previous data analysis. This usually happens when the company is new and there is no previous data to be analyzed, or when a company is dealing with a change of leadership which in return asks for partial, or at times, total reconstruction of how the company works. All of these are daunting tasks that sometimes can be avoided, but other times must be undertaken (Laursen, Thorlund; 2010).

We should have one thing in mind, especially when we deal with a new company which is just entering the world of business. Yes, such a company indeed lacks its own data upon which to make strategic moves, still, there is no obstacle to evaluate how other companies do their stuff; they should learn both from the ones who are successful and from those who have failed. Data Science encapsulates all of this to provide the needed tools to overcome every possible outcome. It is to be believed, of course, that there doesn't exist a single formula of success, but analysis, continuous proficiency and training are the most important aspects where Data Science can help a lot.

One other thing that is crucial in this area, is what data is important and more relevant at what point in time. This is a tricky thing which needs to be considered, because although a specific data seems highly important, and in fact, it may actually be as such, still, that particular data may have no high relevance in a given problematic environment. This means that we should be aware of what is important for a specific issue and waste no time in something which seems important, but in fact, it distracts us from the main thing. At times, it may be needed to sacrifice a given data for the sake of doing the job faster. In other words, data analytics is a complex and sensitive process for which we hope that Data Science will continue to be a solid supporting ground to achieve the strategic points of any company or educational institution to do the right thing at the right time.

When a company develops its strategic plan, it certainly means that it should set specific targets that need to be achieved based on data analysis outcome. The target can be of any nature such as evaluating how new products are selling, how older ones need to be restructured, how human resource needs to be managed, does the company

need additional working force, what production facilities need improvement, and so on. Whatever the target maybe it should fulfil some standards to be considered successful. Those standards are as follows:

The data should be *measurable* because if we face unmeasurable tasks than the target may not be so relevant after all. For example, if we are not sure about the expenses of e given product, we cannot build a decision on such a target; if we're unaware about how many workers we need to achieve a certain amount of producing a given product, the target should be immediately re-evaluated or we should skip to another target.

The vast majority of organizations decide based on mutual coordination of a board of decision-makers, thus if any given decision is to be made, it should be agreed by each one of those who make decisions. If in any case, certain members of the board are against anything, there might be problematic data analysis resulting in unclear reports, which may lead to wrong decisions that may be costly (Laursen, Thorlund; 2010).

Every organization is of different scale and capability, meaning that not all of them can achieve large objectives; not all of them can produce the same number of products. This leads us to the third standard, which is that the targets must be realistic. We cannot expect an organization of 20 people to achieve results as one with 200. We cannot expect an academic institution with just a few professionals to provide same teaching capacity as a large university with a large budget and a much higher number of academics. Data Science should and must provide the tools to extract invaluable information from the data so decision-makers can take proper decisions.

Every day we set a goal to do something, no matter if it is for our personal interest, or for our organization, one thing is for certain, we specify, or at least we try to do that particular task or job in a specific period. We can never start something, especially when we are dealing with business tasks, without specifying the time at which the task should be completed. This is another standard that should be fulfilled so the target can be established. We cannot take upon something and prolong its completion indefinitely (Laursen, Thorlund; 2010).

Data analysis represents a crucial aspect in data management, for which Data Science is at the forefront of developments in providing the means to overcome any possible obstacles we may encounter throughout the process of data management.

Data Science to the rescue

As we mentioned until now, data management is one of the most important aspects of managing an organization. Data scientists are those who should be prepared for any possible scheme of action regarding the acquisition of knowledge out of raw data. Data scientists use various methods and techniques to achieve their goals, but one which is gaining a lot of attention is the agile approach to development. The agile approach is very popular almost in any IT field, be it software testing, software development, etc. This approach is usually based on some principles that should be followed. In our particular case, some researchers suggest these changes that should be undertaken so the agile approach can be successful

- Choosing generalists over specialists
- Preferring small teams over large teams
- Using high-level tools and platforms: cloud computing, distributed systems, and platforms as a service (PaaS)
- Continuous and iterative sharing of intermediate work, even when that work may be incomplete (Journey; 2017).

The agile approach is based on team-work, which means that those involved should comply on certain principles, such as time. If you want to be a part of an agile team you should have the time to do it. If you can't find the time, problems are certain to follow.

Trusting each other is another key point of the agile approach. Team members should base their cooperation on trust. They should believe that every other member of the team is capable for the task he has been given and in that way the probability of success will certainly be much greater.

What benefit can we have from a team that lacks communication? Members of a team need to communicate to explain the tasks, to provide suggestions, to ask for help and so on. Communication is a bridge by which different experts can combine their knowledge to achieve the main objectives of the organization (Jurney; 2017).

Conclusion

We believe that we have properly clarified the importance of data management in every aspect of doing business and how important is the approach displayed by Data Science. We can make a point that Data Science is evolving continuously, finding new ways to help those who need to make decisions for their organization; machine learning combined with Artificial Intelligence will certainly facilitate largely the tools for the data scientist, but let's hope that these developments will be done for the sake of the wellbeing of humanity and not just for the sake of technology. The former will make our lives better, while the latter, well, we're not so sure.

References

- Hoffer, Ramesh, Topi (2011); *Modern Database Management*, Pearson, Boston,
- Jurney, R. (2017); *Agile Data Science*, O'Reilly, Cambridge,
- Laursen, Thorlund, (2010); *Business Analytics for Managers*, Wiley, Hoboken
- Patil, DJ. (2011); *Building Data Science Teams*, O'Reilly, Cambridge
- Peng, R.D., Matsui, E. (2015); *The Art of Data Science*, Leanpub, Victoria
- Vermaulen, 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
- Davidoff, Sh. (2020); *Data Breaches Crisis and Opportunity*, Addison-Wesley, Boston
- Ozdemir, S. (2016); *Principles of Data Science*, Packt Publishing, Birmingham
- Prasad, R., Rohokale, V. (2020); *Cyber Security: The Lifeline of Information and Communication Technology*, Springer, Cham
- Kampakis, S. (2019); *The Decision Maker's Handbook to Data Science*; APRESS, London