Big-Data-Science

Best Practices / Critical Succesfactors

Table of content

1) Management Goals/Big-Data Alignment

First of all its important to know which insights/answers you are seeking. Because then you can be sure you are investing the appropriate amount of time and money in the right place and effectively seek the insights/answers that are needed.
Business alignment is about understanding the business purpose for big-data activities and recognizing the financial value that the activity provides to the organization.

This means first the valuable management goals and questions should be described. Then its possible to produce the insight/answers from the data that are needed to reach the management goals and to answer the management questions.

In this process the focus on management/business goals should not distract to technology goals. Its common data-scientists/technologists get distracted by new cool technology and models. However Big-Data should always serve financial/management goals and not the technology goals. Unless offcourse new technology is carefully selected to reach specific financial/management goals. In the Big-Data Journey start with clear management/business goals and gathering and the business requirements.
Describing clear goals and business requirements should always be the first step before beginning the process of leveraging Big-Data Analytics, because it sets the aim in the right/successful direction.

2) Starting with the right questions. And keep asking the right questions

In practice, instead of focusing on how much money a project will earn/save the company its common technologists focus on how cool/innovative projects are. This bottom-up approach (from technologists to management) can lead to technology output that doesn't hit the management goals and doesn't answer the right management questions. Therefore the right management questions should always be stated top-down from management to the data-scientists. However sometimes experienced data-scientists can come up with management questions management can't even think off. Therefore stating the right management questions should always be a team effort with managers and data-scientists involved. Also keep in mind that further in the process there will be new information, causing management questions to change. Therefore close continuous communication between the data-scientists and management regarding management goals and questions is very important.

3) Thinking strategically, acting tactically

When organisations start with Big-Data, they often start with one specific business problem and built a specific solution to solve the problem. Often this first initiative is viewed as an experiment. If successful, then we'll think of more Big-Data solutions. However, in practice first applications typically start slow with lots of organisation specific start-up problems to solve. Therefore this first big-data initiative should be part of a higher Big-data strategy. With such a strategy in place there has been thought about the first application, however there has also been thought about the following second, third and fourth application. Probably in

formulating the strategy certain expected financial numbers are tied to the different applications in different stages. This is important because in practice often not a single, but continued big-data applications lead to significant big-data value.

4) Data quality

Data Quality is crucial for data-scientists. Often data quality is described as 'good', however what does 'good' mean? For example a Sales manager wants a trending analysis for the number and types of products sold in a certain area and for this purpose he needs the postcode to be accurate. However the logistic department needs the street address and number for delivering the products. In this example good data quality is good for the Sales unit, however not so good for the logistic unit.

There are several dimensions to good data quality including timeliness, completeness, validity, accuracy, consistency and integrity.

The right way is to clearly define the desired data quality dimensions and requirements for each stated management/business goal. Next the data understanding(metadata) should tell you if the data meets those requirements. If it's not the case, it should be clear who to contact for the steps needed to improve the data-quality so the data-scientist has the resources he needs to reach the goals he has to reach.

5) Data-centric processes

In data-centric processes, data understanding is gathered as new data is created, and data quality is measured when that new data is monitored. Data-centric processes occur in many departments where data is viewed as an important part of the process and not just as output of the process. A project management process can be data-centric when the data needed is the input to the definition/design of the project. A systems development process can be data-centric when the data is not just considered the functionality, but also the UI, infrastructure and hardware.

Operational processes can be data-centric when it's equally meaningful to create new data in an accurate standard way, as it's meaningful to execute the process within a specific time frame. When an organization implements data-centric processes, the data is of higher quality and better understood. In Big-data driven companies this is important, because the data-centric processes reduce the huge complexity of big-data to streamlined data/information streams enabling employees to efficiently handle the big-data. These data-centric processes also enable more effective, agile and real-time analytics.

6) Support from the company culture

Support from the corporate culture is very important, because in practice the company culture accepts/supports the big-data initiative or it kills the initiative. In practice for most departments who are not used to the data-driven way of working, a data-driven culture can

sound difficult and intimidating. A new data-driven way of working is then viewed as a threat for their job safety and therefore there will be resistance against a new data-driven culture. Most employees are against any change at all let alone disruptive big-data changes.
The best way to get support for big-data initiatives in the company culture is to seek support in the top. Negotiate with high level stakeholders/sponsors and understand their needs. Work with the existing reports first to get informed on the way data and analytics is used in the company now. Then extend the existing practices with big-data analytics reports including new code, formula's and data-sources. This way of working leads to employees understanding the starting context of the big-data initiative(because its their old way of working.) For example Excel, PowerPivot and PowerBI reports can be the starting point. Then from there more advanced big-data initiatives can be extended to the old way of working and the employees can progress with you in the new big-data journey.

7) Involving all Business Units in the Big-Data Strategy

Big-Data should never be an isolated activity within the Organisation. Far more its a central function and data can be viewed as the life-blood of an organization. Without data there's no information/knowledge transfer in the organisation and the organisation would freeze. With Big-Data, organisations can leverage huge volumes of data to learn more about customers, processes and events than would be possible with snapshots of data. A properly executed big-data strategy can have a large impact on the effectiveness of business functions. Big-Data can add different types of new external data-sources. These out-of-the box data-sources can provide surprising new perspectives, valuable new insights and bring significant business value across business functions. Often its the holistic or 360-degree data-driven view that leads to very valuable insight.

Therefore collaboration and communication across departments is very important (*Please also view our article the big-data-science process*). Its also important for data-scientists to understand what the business/managers are trying to achieve. And business/managers should at least have a high level understanding of what the business can and can't achieve with big-data. This is important for setting the right frame in the big-data-science process setting and keep setting the right goals, keep asking the right questions and stating the right hypotheses. This is something that can only be done effectively if the managers/business and the data-scientists understand each other and can communicate effectively.

8) Investigating what you have and what you need in Big-Data

A lot of data is not equal to good useful data. The strategy of first collect all data and sort it out later will most likely fail. That is because the more roughly data is collected, the more disorganized the data becomes in varying formats. It will then get harder and harder to find the right useful data in this disorganized bulk of big-data. Therefore its much more effective to be goal-oriented. Which data do we have and is useful for a specific goal? And which data don't we have, but should we have for this goal? Make sure you have all the data that is needed in the right place, before starting the project. Offcourse its not always possible to know exactly which data you need beforehand, so keep the flexibility/possibility to keep adding new data to projects/models during the data-science process. Most important here is to really have clear which data is useful and which data isn't. Therefore quickly test data and

select valuable useful datatables, sets and variables, before starting the project. Filter out the useless data. Especially if you come to work with other data specialists its unprofessional to bring in unclear/messy data into the project that isn't useful at all. So the data should be prepped for a clear/effective start. Have the right data in place before starting the project.

9) A stepwise approach

Its a common mistake to start tackling a new problem with new/unknown data. The most successful organizations start with solving a known problem in a new data-driven way (A new tool for example). Then they solve the same type of problems with new/unknown data and then they move on to solving new problems with new data. Organisations that take a stepwise approach have a much higher chance of success. The art of reducing/demystifying the huge complexity in big-data-science to clear effective valuable actions and describing them effectively in the right order of priority in a stepwise implementation plan may even be key to success in big-data.

10) Describing the big-data road-map

At a certain stage, there is experimented with data, the companies goals and objectives are clear. There is a good understanding of what the higher management and business units need. Then its time to create the big-data road-map. As stated earlier the big-data roadmap or implementation plan is key/critical for big-data success. The road-map is essential for several reasons. First, its a clear guide that can be followed for future success, second, It gives you a tangible document with a detailed process. Often roadmaps/process maps become increasingly more complicated/detailed with each new version. There is no way you can grasp the complexity/details of the 21th version without a tangible process-map document. Third, its the central big-data process everyone in the organisation can view and will follow. It makes sure there is a unified view on the big-data initiative and the big-data actions. Every person in the organisation has the same clear mental picture when it concerns the big-data initiative. This is very important.

11) Continuous communication

Continuous communication between stakeholders and data-science is essential for big-data success. For example management goals and questions change during big-data projects and should change, because as new information is available better goals and questions can be stated. So when new information becomes available during big-data projects, management should communicate this effectively to the data-scientists and vice versa to be able to quickly asses if and how the new knowledge effects the project frame(Aim, goals, questions etc). Also to do this effectively the data-scientists should understand the management perspective and the management should understand the data-scientists perspective. Therefore close communication is essential for a smooth growing relation. So continuous communication is essential for keeping the right frame(Aim, goals, questions etc) when new information becomes available.
The bottom line is ,stating the right frame with the right aim, goals and questions etc. is an art and is half of big-data success. The big-data operations can be perfect, but it they answer

the wrong questions, then the operations and results are useless. Thus stating the right frame for big-data projects is an art and is something Management and the Data-Scientists should do and get better at together.