KISS

MY DATA

THE 7 TEMPTING PROMISES OF BIG DATA

Big Data continues to create buzz and feed the wildest expectations within organizations. Without objectification, some of these expectations will remain at the stage of fantasy and will so inevitably frustrate and disappoint. Let's get real.

1 | TAKE ADVANTAGE OF GIANT VOLUMES OF DATA

It is almost a pleonasm, Big Data allows the exploitation of massive volumes of data. Yes it's also one of the constituent component of its definition based on the 3 V's (Volume, Variety and Velocity). The amount of data managed in Big Data bases are theoretically infinite (through an extreme scalability enabled by these architectures). However, most real projects usually oscillate between a few hundred gigabytes and tens of terabytes of stored data.

We regularly get asked the question as to whether it is relevant to carry out a Big Data project knowing that it will not exceed volumes beyond terabytes. The answer is clear: yes! The amount of data to be processed does not constitute the first criterion to justify the use of Big Data. There are 6 more attractions you should take into consideration.

Moreover even if the current volume of data you wish to exploit is not that vast (yet), you have no guarantee this will not evolve in the future. Indeed, once the first initiatives are in place, you will be tempted to exploit new information, which can potentially be huge. In addition, with the new and fast spreading use of sensors and other connected objects, your company will more likely have to deal with higher volumes of data, which can also quite rapidly exceed terabytes.

2 | MAKE USE OF NEW SOURCES OF INFORMATION

Gartner measured that only 30% of corporate data is used. This means that there is 70% of the information being left undiscovered, opening up not yet activated opportunities. This internal information includes such things as contracts, reports, customer-service exchanges, etc... but also Excel tables scattered over the company workstations.

In addition, it is the exploitation of external data that is source of value:

- Open-data or data resold by other companies (INSEE, weather, telecom operators ...) to contextualize and enrich the understanding of your environment.
- Sensor data or data from connected objects to acquire daily information, about the place of use of the studied subject (factory machine, a SI log, the way a client uses a product...)

All these data, that are still underexploited in most cases, open new fields of investigation and will certainly generate new business opportunities.

3 | PLAY WITH UNSTRUCTURED DATA

Until now, only the data known as "structured" were utilized by organizations. Structured means that it must be possible to organize information in pre-determined boxes (imposed by the SQL database models) and that this information can be easily manipulated (as it is the case for 90% of digital data). Gradually, this has led our organizations to wear blinders and to ignore the wealth of knowledge that other types of data can reveal.

With Big Data, everything has a value, a meaning. This way, one can exploit information which structure is yet unknown or which shape are likely to develop. One can also use images, sound and video. For example, the French National Railway, SNCF, analyses photos of rails (taken by trains while in motion) in order to identify the potential locations where preventive maintenance needs to be scheduled.

4 | PREDICT
THE FUTURE

Statistics is not a new discipline. But it's known a new upswing as a result of the innovative Big Data interest. We now talk about Data Science, Data Mining and Machine Learning techniques. Their purpose is to better understand the current situation (descriptive statistics) and to predict the future as precisely as possible.

Conducted by algorithms, this intelligence becomes necessary because of the downpour of data, which can no longer be manipulated on a human scale. Also, it is necessary to fully exploit the large diversity of the "Big Data" data – particularly if you want to profit from temptations 1 to 3.

The most advanced usages of Data Science techniques already allow you to predict the future. On a recent project, we were able to predict the turnover of various facilities of a well-known catering chain.

These new Data Science applications create numerous fresh opportunities and provide remarkable new insights, especially when executed in real-time.

5 | INTERACT IN REAL-TIME

This is certainly one of the most mind-blowing applications in terms of Big Data and Digital Transformation. This is because of the ability to ingest and process data in real-time on all kind of different digital supports.

As a result, this opens up an infinite application scope: websites adaptations in real-time according to the navigational behavior and/or the knowledge about the visitor, push promotions targeted on the best channel, interactions with the visitor at the physical point of sale, etc.

From a technical point of view, technologies such as Spark or Flink and architectures like Lambda already deliver an open-source answer for this temptation.

6 | BRING INTELLIGENCE TO THE HEART OF BUSINESS PROCESSES

This is perhaps the most exciting Big Data temptation. Big Data allows the data to be placed at the heart of all business processes; we then talk about micro data-driven decisions. This is what large web actors like Amazon do when optimizing every move of its operators within the supply chain.

The previous temptations have demonstrated that Big Data and Data Science have helped create intelligence and bring reactivity to decisions. Like the cherry on the cake, Big Data technologies have a native capacity to integrate operational information systems (ERM, CRM, etc.). Indeed, these technologies come from the web sphere, and quite often use the Java language. Interfacing via web services on SOA enterprise architectures are in fact easy to achieve.

7 | PAY LESS

An already common definition states that "Big Data is a supercomputer within the reach of all." What makes this especially exciting is the second part of the phrase: "within the reach of all".

The structural operation of Big Data technologies is to be based on ordinary hardware architecture (x86 machines) and an open-source OS (usually Linux). This helps drastically reducing the infrastructure cost. Moreover, many software solutions are open-source (including the inevitable Hadoop) and thus make it possible to drastically reduce the costs of licenses as well.

This lower cost is often recognized as a primary motivation for using Big Data technologies. Without even implementing something new, we can do former actions at a reduced cost (including for storage and calculations).

How tempting is this…? So why resist?

Business
& Decision

ROI²

return on investment
through
return on information

If you want to know how Business & Decision can help you
face these challenges, please contact:

ARNAUD PICCIN
Big Data Practice Manager
Arnaud.Piccin@businessdecision.com