Big Data Can Rock Your Business (If You Let It)

Martin Triska

Big Data Technology Enthusiast, Ataccama

Big Data is all around us. It is not likely that you heard this phrase for the first time. It is, however, very likely that you still do not understand what it really means and—more precisely—what it means for you. Simply said, Big Data represents a new paradigm of data gathering, storing and processing, which can revolutionize or at least significantly improve many of the typical data-focused tasks, such as data integration, transformation or analysis. Incorporating a Big Data platform into your infrastructure also enables you to do things that were not possible before: storing all information together, focusing on new data sources and types (mainly unstructured data as text, voice or video) or simply scaling the environment easily.

Big Data is not a simple one-size-fits-all magic box you could buy. The suitable solution heavily depends on your specific needs and requirements. While you think of your own scenarios, take a minute and read the following customer stories. Maybe some of them will inspire you.

Enterprise Data Hub: 360° Customer Intelligence

Understanding your customers well is usually the single most important challenge any business faces. It is not just about what to sell to whom, but also when and— with increasing importance—how. To answer these questions correctly you need not only smart algorithms but also the right data. Luckily for you, the customer is willing to give you all necessary inputs for free—you simply need to listen, just like the following European bank with their Big Data solution.

Their goal is rather simple: to collect all information customers are willing to share, integrate it into one place and, for each person, extract the important characteristic. Did he mention he is thinking of buying a house? Will she go abroad next month? Would they be good candidates for travel insurance? What about their kids? Do they even have any?

The beauty of storing all the information permanently is that it may hold answers to the questions you might not have even thought about yet. Data like complete customer communication (emails, social media activity, call center transcripts, etc.), notes from the interaction between the customer and the company (notes from clerks, field workers, etc.), every single activity the customer has ever done (logging into the web application, withdrawing money from an ATM, paying with a credit card, etc.), and more. Thanks to the Big Data platform, it is possible to query all this at once, correlate it with other customers, and instantly gain insights into interesting patterns of customer behavior.

Real-time Traffic Optimization

Not every Big Data problem needs to involve large volumes of data. Sometimes an innovative idea might come just from handling the incoming data stream fast (or better, in real time). For example, the City of Cape Town in South Africa optimizes its public transportation system according to the current real load, i.e., by tracking the number of actual passengers, where each passenger is represented by a mobile device. When there are too many mobile devices (too many people) at the same line, it is reinforced by rerouting of a less utilized vehicle.

Getting Under the Big Data Hood

No matter what case you choose to start with, you will soon begin wondering how Big Data works under the hood. It is no secret that the heart of most solutions is powered by a technology called *Hadoop* and applications from its ecosystem. Be it *Hive*, *HBase* or *Flume*, every single piece has its important role in this puzzle which cannot be easily replaced by other technology. Although it neatly fits together from the architectural point of view, this fragmentation imposes immense requirements on knowledge and skills required to start even with small steps. Usually people do not want to learn completely new technologies

just to be able to leverage their benefits, and this is especially true for the *Hadoop* environment. For example, it is hard enough to master *MapReduce* (parallel computational model on *Hadoop* clusters) to receive the expected results, and less straightforward tasks are even harder.

Another problem you will eventually face when you let more data into your environment will be data quality. You will suddenly lose control over the location or consistency of your information; unstructured data will make data quality efforts even more challenging. It is unrealistic to expect that a mere act of putting the data together will instantly lead to better results from data analysis. The idea is to leverage big but also clean data. The best scenario is to use your master data to drive the Big Data initiative to gain the best possible results. It is important to clean the data before you store it in *Hadoop* and associate/create the correct metadata. The more data you have, the more important it is to understand what it consists of. This leads to a need for some "data discovery" solution. It is usually an iterative process of discovery where problem definition and algorithm design repeat itself until the business expectations are satisfied.

It is, therefore, very important to make it as easy as possible for all users—not just the technical staff in your organization—so they are able to use all Big Data tools easily and literally abstract the complexity away. Many vendors these days start to offer connectors to *Hadoop* environments from their mining/analytical software. This is not the solution, as all they do is bringing all the data required for the computation to one place over the network. You must, however, seek a solution that will enable you to run jobs directly against the cluster—seamlessly and in a manner that is possible to debug, test, and document.

Luckily, Ataccama offers an easy way to leverage the benefits of Big Data tools without heavy development or steep learning curve: *Ataccama Big Data Engine*.

Big Data Test Drive

Ataccama Big Data Engine represents an easy-to-use development interface (GUI), shared metadata, and rich data integration and processing capabilities—replacing specialized ETL technologies. It accommodates the key Big Data features, such as massive parallel processing, scale-out nature, fault tolerance, memory management, and is available to anyone who needs to profile, map, model, process, transform, cleanse, enrich, and integrate Big Data.

Ataccama Corporation combines data quality, master data management, and data governance in a single technology platform ready for operational, analytical and BigData deployments. Kick off your Big Data initiative today and request a complimentary *Big Data Test Drive*. Our team of Big Data Enthusiasts will help you with a *Hadoop* cluster setup, data assessment, assessment of your potential Big Data ideas, and, finally, also with one Big Data idea implementation.

Just as we like to say—Big Data is not a simple one-size-fits-all magic box you could buy. Ataccama Big Data team therefore listens carefully to discover your needs first. Then, we deliver a suitable solution matching these needs. We simply work hard to make your work with Big Data easy.

. . . .

About the Author

Martin Triska is a Big Data Technology Enthusiast at Ataccama. Prior to joining the company, Martin had been working in the field of Big Data technologies for a few years. Thus, he possesses substantial background in this area as well as in text analytics. He gained his expertise from both business and research practice. Martin holds a Master's degree in Software Engineering.