

[Home](#) > [AI/ML](#)

The Importance of Data Culture in the Data-Driven Business Era

7 June 2024 - 6 min. read

[AI](#) [Data and Analytics](#) [Data Ingestion](#) [Data Security and Governance](#)

[Data Visualization](#)

In today's digital world, data plays a crucial role.

It enables companies to stay competitive by responding quickly to the evolution of their target markets through a faster decision-making process. This results in better business strategies that are increasingly focused on improving user experience.

However, being "data-driven" only happens if the company has a deep and shared Data Culture.

Effectively instilling a Data Culture within a company is a slow and complex process that must involve all organizational stakeholders, from management to decision-makers and IT experts such as data analysts.

But what is Data Culture in the data-driven era? And why is it essential to successfully transition from a more traditional approach to a data-driven one?

In this article, we will dive deep into this topic.

Information VS Data

Let's start with a definition to gain a correct perspective immediately.

To make better decisions, we need information to which we can attribute meaning. Information doesn't innately exist; it must be extracted by processing and interpreting

data, which are the building blocks that compose it.

Therefore, **"data" and "information" are not precisely synonyms**: not all data can become information, as we will see later in the article, but having data is a necessary condition to deduce information impacting decisions.

Data Sourcing

With this premise, it is clear that the first thing a company must do is be able to collect data.

Obviously, it is unthinkable—and not useful—to collect everything right from day one.

Our advice is to start with a detailed mapping of everything that can generate data within your organization (for example, applications, machinery, probes, sensors, and, in general, user interactions) and then structure the data collection accordingly.

How much data do I need for a successful data-driven project?

In the data world, it is impossible to visualize all the possible scenarios of a project from day one. Therefore, an abundance of data is often an advantage.

However, data can acquire value only if historicized, put in relation with each other, and archived within a single Data Lake containing a sufficiently large dataset. From this point of view, the Cloud offers numerous possibilities. In particular, storage tiering logic allows for the cost-effective archiving of vast amounts of data for short-term decisions and, more importantly, for future competitive advantage and faster go-to-market.

However, owning a lot of data without worrying about the cost is not enough to ensure a (good) result. There are countless variables at play, especially when— as it should always be—the goal is to use data to do something new and innovative. For example, during the project, it is common to realize that the initial issue we wanted to address using data is not the right one in relation to the data we actually have or that the answer obtained from the collected information is not what we expected.

To explore data and understand its potential, it is recommended that you conduct an **Exploratory Data Analysis (EDA)**. This allows you to visualize graphs and descriptive statistics about data's characteristics and to identify **patterns, anomalies, and relationships** between variables.

Tools like Python and R, with libraries such as Pandas, NumPy, Matplotlib, and Seaborn, are essential for performing EDA efficiently.

Additionally, tools that assist Data Analysts, like Jupyter Notebooks, can make the analysis fast and effective.

Another fundamental technique is data mining, which uses advanced algorithms to discover hidden patterns in the data. Tools like Apache Spark, Hadoop, and SAS are used to process large volumes of data and apply clustering, classification, and association algorithms.

Our advice is always to maintain a **short feedback loop** between when the data collection begins and when the results of their processing are analyzed. Frequently ask yourself:

“What story is my data telling?”

This way, you can identify strong business indicators you still need to consider, which could lead to advantageous **adjustments in the overall commercial strategy**. This iterative and incremental approach goes hand in hand with the fail-fast principle typical of the Cloud. The low cost of failure is an incredibly important innovation boost since it allows companies to experiment sustainably, explore multiple paths (even in parallel), achieve results quickly, and significantly control risks.

But be careful: too much information means no information at all!

It's all about identifying the right business objectives and collecting the relevant data to achieve them accordingly. This part comes with many uncertainties; the support of a trusted technology partner with deep experience in Cloud and end-to-end data projects can really help you build a good data strategy.

Data Governance, Security and Compliance

Although it seems like a simple and somewhat "obvious" step, one of the biggest challenges for organizations is creating a Data Platform to make data accessible to multiple professionals securely and efficiently.

Once we have collected a certain amount of useful data, it is time to use it to extract value. This is when governance, compliance, and security issues arise. We need to start managing permissions correctly to ensure the right information gets to the right

person at the right time. We must define access rules, identify the entities that can operate on the data and the rules according to which they can do so, and categorize all information as public or sensitive data. There are various strategies for this. We recommend cataloging all data as "sensitive" for less complex situations, possibly already masked and anonymized. This way, you'll have fewer authorized roles accessing the data, thereby significantly reducing security, control, and compliance risks.

Instead, in larger organizations where many stakeholders can benefit from data usage, considering everything as "sensitive" can be limiting. One of the best strategies in this case is to label the data at the collection stage, classifying it correctly based on its level of confidentiality.

Another recommended best practice for managing and securely handling a large amount of data is the use of multi-factor authentication mechanisms and role-based access control (RBAC) policies, which clearly define each user's permissions based on their responsibilities.

Another crucial aspect is ensuring data quality by defining data verification and cleaning policies. This involves creating standardized data integration, transformation, and validation procedures, ensuring that the data lake's information is accurate, complete, and reliable.

Adopting monitoring tools that can help detect suspicious activities and mitigate associated risks is useful for maintaining data quality and security.

Of course, once we have categorized the data, we also need an efficient way to query it so that the right people can access the relevant data.

To Conclude

In this article, we delved into the crucial importance of **valuing data, making it accessible, and integrating it into the business decision-making process**. All these actions are fundamental for transforming information into a true competitive resource.

We highlighted how being aware of the value of data is the first step in managing it appropriately in order to extract value in the future. Promoting a data culture requires not only understanding and adopting new technologies but also transforming how the entire organization perceives and uses data. The whole company must be involved in

this collective effort, from top management to individual employees, to overcome existing barriers and fully integrate the value of data into daily practices.

In conclusion, adopting a Data Culture requires a holistic approach combining data governance, continuous personnel training and education, and implementing appropriate strategies and tools.

Only through this strategic and collective commitment can companies transform data into a sustainable competitive advantage, successfully addressing the challenges of cultural and technological change.

About Proud2beCloud

Proud2beCloud is a blog by **beSharp**, an Italian APN Premier Consulting Partner expert in designing, implementing, and managing complex Cloud infrastructures and advanced services on AWS. Before being writers, we are Cloud Experts working daily with AWS services since 2007. We are hungry readers, innovative builders, and gem-seekers. On Proud2beCloud, we regularly share our best AWS pro tips, configuration insights, in-depth news, tips&tricks, how-tos, and many other resources. Take part in the discussion!



Alessio Gandini

Cloud-native Development Line Manager @ beSharp, DevOps Engineer and AWS expert. Since I was still in the Alfa version, I'm a computer geek, a computer science-addicted, and passionate about electronics all-around. At this moment, I'm hanging out in the IoT world, also exploring the voice user experience, trying to do amazing (IoT) Things. Passionate about cinema and great TV series consumer, Sunday videogamer