
Six important trends of the 2022 data and analytics landscape



CEO, Marc Delbaere

“ One of the nice perks of scaling up a data software business (besides the sleepless nights) is that you end up meeting so many interesting people. Customer interactions with practitioners or decision makers are already providing a very wide spectrum of perspectives and when you add to that integration and technology partners, industry analysts and venture capitalists, you can accumulate an incredible amount of information very fast .

The focus on developing your business gives you the right lenses for meaningfully structuring the conversations and little by little you are starting to discern the figure in the carpet.

I thought I would pause for a second and share with you what I see right now as some of the most defining trends of the data and analytics market. ”

01. Skills shortage in data engineering

As every company is struggling to deliver its data projects, the war for talent is in full force. Yet, despite a very competitive job market, data engineers are still spending a lot of their time on repetitive tasks. While the data engineers are busy fixing unmanaged point-to-point data pipelines, the data scientists are waiting without the input required to do their work. In some cases, the time spent on data engineering can go up to 70% of the overall workload of data projects.

Solutions automating tedious tasks and offering self-service access to less technical profiles are in very high demand. Applying these principles to the data engineering domain, the demand is high for no-code ingestion (connecting to data sources through pure configuration), low-code transformation (visually or through using widely available skills like SQL, which remains more than ever the lingua franca for data manipulation and transformation) and no-code distribution.

02. Common approaches across analytics and AI/ML

As the needs for leveraging data are increasing across the board, leading companies are starting to realize that managing two different data infrastructures for business intelligence and AI/ML is not sustainable in the long run and are extremely interested in approaches that can efficiently serve both needs. The rise of architectures like data hubs or data meshes is largely a consequence of the limitations of data warehouses and data lakes to serve this dual purpose.

While rationalizing data infrastructure seems to be the logical thing to do, some architectures have limitations that can limit the long-term possibilities. The risk is to standardize on a data stack and apply it to use-cases for which it was not designed. Some of the most common shortcomings are data quality issues, lack of flexibility for the intended purpose or high latency.

03. The rise of data in motion

As businesses are becoming more digital, want to provide the best customer experience, and want to react instantaneously to incidents, they need to update their data foundation to provide actionable, real-time insight.

Machine learning applications and systems depend on massive and continuous volumes of data being processed and analyzed in real-time. At the same time, according to IDC, data generated from connected IoT devices are projected to be 73.1 zettabytes by 2025, almost four times the 18.3 zettabytes generated in 2019. IoT data is best treated using streaming technology so that it can be processed in real-time.

Most data infrastructures are not calibrated for real-time, and customers are looking at lower-latency solutions. Streaming technology is now ready for prime time but remains hard to deploy particularly under the skills shortage described earlier.

With a real-time foundation, you are also solving another big problem: the continuous reconciliation processes of disconnected data sources. By consuming data in streams, you are always using the latest version of the data.

04. The need to support hybrid and multi-cloud data management

Some enterprise applications like CRMs or HR systems are now routinely cloud-based. Many external sources of data available through APIs are used to improve business intelligence or help train machine learning models. Yet, many data sources used and produced by legacy systems are still managed on premises.

The need to reconcile data across these different environments is higher than ever. Cloud infrastructure providers are fiercely competing to provide the best single environment from which enterprises could manage everything, but reasonable risk management practices are pushing enterprises to hedge their bets to avoid cloud vendor lock-in. If you add to that the geopolitical factors associated with physical data storage, it becomes very clear that hybrid and multi-cloud capabilities are becoming a must.

Gartner estimates that nearly half of data management implementations use both on-premises and cloud environments. Data integration tools are expected to dynamically construct or reconstruct integration infrastructure across a hybrid data management environment.

05. Retaining control in decentralized environments

As data organizations are evolving towards more decentralized architectures, with more empowering of end-users to get rid of IT bottlenecks, the risk of mismanagement is increasing significantly.

You want to avoid that sensitive data falls into the wrong hands or is used for the wrong purpose, especially when you are dealing with personal data under strict regulation like GDPR in Europe.

The tension is real between the need for self-service data infrastructures and the absolute necessity to govern the data end-to-end. The best of both worlds consists of infrastructures that can serve data very easily to end-users while keeping a full lineage of the data and enforcing privacy requirements.

06. The need for speed

The pressure on the business to deliver additional results through clever harvesting of its data resources has never been so high.

On the top line, between the improvements in customer satisfaction resulting in higher retention rates, churn detection and prevention measures, or higher revenue per customer through better recommendations, there are many data-related opportunities to beat competition.

On the cost side, with an overall inflationary context pressuring the margins, optimisation of production lines, energy cost reduction, procurement cost cuts are some of the numerous areas where data can make a difference.

Once you consider the overall bottom-line value of successfully executing all these data-related opportunities, the difference in execution speed of data projects becomes a primary competitive advantage.

If you take into consideration that the pressure on data infrastructures is compounding when you increase the number of data projects, getting a fast path to value while keeping everything under control is becoming a strong business imperative.

While there are certainly other important trends that I am not mentioning here, I can testify of the important number of conversations that I have had in the last couple of months that are highlighting the six trends that I have selected for this post.

To find out more about the world of data and analytics, visit our website www.digazu.com.