



The answer to  
your business  
problems isn't  
more data,  
it's 'smart' data



Organisations of all sizes, in all sectors, are drowning in a sea of data. After over a decade of exponential growth of big data concepts and mighty lakes, are we finally reaching the point where the tide turns and we ask ourselves if less is more?



## Upping the data ante

The value of data is undeniable, and the advantages of big data are widely understood. However, the reality is that many organisations now struggle to wield the amount of data they already have. And yet these organisations continue to collect as much data and information as possible, in any format, and any level of quality, without really processing or structuring it in any usable way. They claim to be data-driven and intelligence-led, but are they really?

This trend started 10 years ago, with the introduction of new data concepts and technologies, in particular data lakes, promising to solve the enterprise's complaints about information silos, undermined productivity, hindered insights, and poor collaboration. Capable of holding vast amounts of raw data until needed, for any use across the enterprise (even for use cases not yet identified) and limited only by storage and compute capacities, these repositories were supposed to be the dream solution, able to combine structured and unstructured data into a one-stop-shop for business insights, integration, and even complex analytics.

Yet 10 years later, many of the challenges described above still persist. We have all undoubtedly been overexposed to the 'data lake vs data swamp' analogy. We could possibly explore whether data lakes are delivering on the promise, or whether data warehouses are better. (BTW, these two ideas are in fact complementary, not an either-or choice). The real question is whether the current 'let's collect and keep everything' approach to data and information lifecycle management is providing tangible benefits or becoming a burden on the enterprise?

## More data does not mean more business agility

Organisations are only as agile as their data; data is the lifeblood of business. But to truly become data-driven, they must also make effective use of their data, and manage this asset through its lifecycle, extracting value at every stage.

Our modern technology landscape is both a blessing and a curse for organisations that aim to become data-driven and subsequently data-centric. It's a blessing, because the significant improvements in cloud computing systems are virtually enabling unlimited storage and computing power on-demand to run state-of-the-art data repositories, analysis engines, machine learning and AI systems. But it is equally a curse, because the same powerful and flexible technology solutions perpetuate the illusion that by simply collecting and storing the data you can magically create true value from the assets. This is not the case.

Current thinking says that to become more data-driven you should collect vast amounts of data and information while leveraging smart automation solutions, machine learning, and artificial intelligence systems. Unfortunately, what this type of approach critically underestimates is the importance of well-designed data management systems, effective data governance principles, and a comprehensive approach to lifecycle management. We shouldn't treat data stores like a vortex, throwing everything into it, forever, and hoping for the best. Data must be carefully managed, curated, protected, and respected to extract its real value.

But due to the sheer volumes of data, many organisations can't (and shouldn't) try to govern all of their data the same way. According to a recent Splunk Inc study on the state of 'dark' data, it is estimated that over 75% of an organisations' data is 'dark' – unknown, unused, and untapped – and this number is growing at an alarming rate. Data stores are like a giant iceberg where only the tip is visible to the organisation. In this context, perhaps we should focus on 'fast' data and 'smart' data rather than just more data.

# Building minimum viable data

The strategic opportunity is for organisations to realise that while complex initiatives to clean up the existing bad data in their estate may be of value, a better approach is to clean up only the most critical data – the Minimal Viable Data (MVD) for which there are known and achievable use cases.

Similarly they should also focus on improving the way new data is created and ingested, including eliminating the root causes of previous errors or limitations to ensure they don't repeat the same mistakes. The MVD approach focusses your team's efforts on the data which matters and the insights which will deliver the greatest value.

Extending on the MVD approach, our obsession for traditional ingestion, ETL, and data modelling should be converted to building flexible and innovative data logistic pipelines that support organisations to integrate data on-demand. This allows us to gain more value from the data now while building rapid capability to respond to business changes in the future. Our efforts should also be directed at the contextualisation and harmonisation of the data to create feature-rich repositories that can integrate easily in the enterprise and provide immediate business value.

The concept of a Minimal Viable Data approach is the only sensible way to manage and leverage our rapidly increasing data stores. Here are the six basic steps to get there:

- 1** Agree on what data is important, what is really needed, and why.
- 2** Decide how much is just enough to make the right decision.
- 3** Understand where the data resides, as well as its relevance, quality, and lineage.
- 4** Create the 'right' governance structures, including a policy framework and roles.
- 5** Design and build the data system to ingest the data respecting the points above.
- 6** Consistently iterate and re-evaluate the value of the system to keep it relevant.

Becoming data-driven is a long and complicated journey. The speed at which technology continues to evolve, coupled with the overwhelming amount of data and information available, makes us wonder if we will ever keep up. Just remember too much data can be as dangerous as too little. A MVD approach will help to leverage the true value of your data and keep your data 'iceberg' from melting.