Informatica | Capgemini

How AI is bringing the data-powered organization to life

Making modern data management the backbone of data estate modernization

How AI is bringing the data-powered organization to life

The world has changed. Economies shuttered by the pandemic are now trying to restart. Those who've been able to keep operating are working in different ways.

Organizations that followed a cloud-first strategy may have been more resilient, but what about now? Without a data-powered plan, will they be able to empower their employees with the information, applications, and analytics to reinvigorate productivity and competitiveness?

Change will require the right, trusted data to flow to the individuals, teams, and information supply chains that energize research and development, enhance the skills base of the current and future workforce, and enable new services.

And change is needed. Despite talk of an imminent fourth industrial revolution driven by digital transformation, productivity was sputtering even before the pandemic — and in some cases stalling.

In the decade between 2010 and 2019, the UK recorded its lowest productivity growth in two centuries — 0.3 per cent.[1]

How can that be—when we are continually being told about the speed of transformation around us? Economist and futurist Tim Harford[2] has highlighted how little has actually changed in the last fifty years compared to the almost non-stop arrival of labor-saving innovations between 1920 and 1970.

"Given how much we keep being told about the disruptive pace of innovation and the boundless creativity of Silicon Valley, the reality is both surprising and disappointing."

– Tim Harford, Economist

From the arrival of e-commerce through the launch of the iPhone and the ongoing migration to the cloud, technology's ultimate promise has been articulated again and again: it will accelerate productivity gains, free people from paper and process, and make it easier for businesses to give customers more of what they want — in all the ways that they want it.

But numbers don't lie. Despite everything futurists and policymakers have told us about the next shiny technology's power to transform, reality has often fallen short of the promise.

Something's stalled in our productivity engine. It may be clichéd, but if data is the oil that keeps the IT cylinders turning, too many organizations are down a pint. To power productivity, data has to flow freely – transformed to be secure, connected, trusted, ethically managed, and democratized by design.

When it comes to data, everything has changed—and nothing has changed

Information generation and capture is an area where we've well-exceeded expectations. The volume of data on planet Earth will hit 44 zettabytes this year — 40 times more bytes than there are stars in the observable universe.[3]

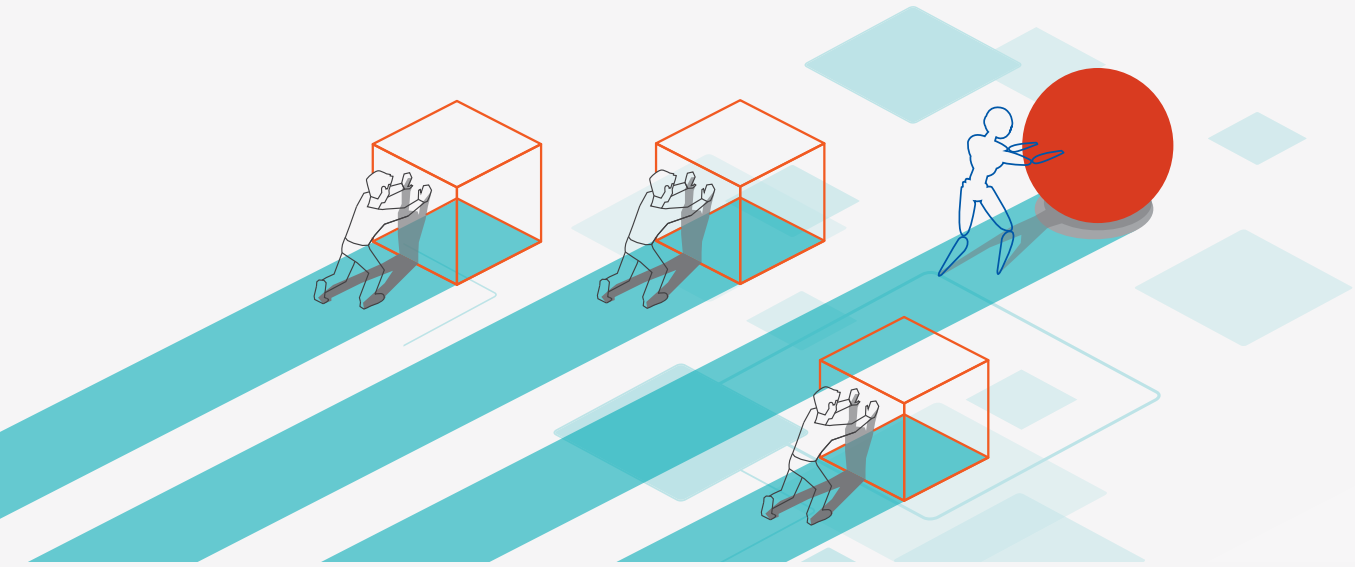
But despite our ability to create data, specialists are still seeking productive ways to reliably, consistently, and efficiently make the right information available, trusted, and secure for the people who need it — when they need it.

Many organisations still rely on making decisions based on information that is at best dated, and at worst is unreliable or inconsistent. This rear-view mirror approach to analysis and decision making won't allow organizations to accelerate with confidence or create the value data promises.

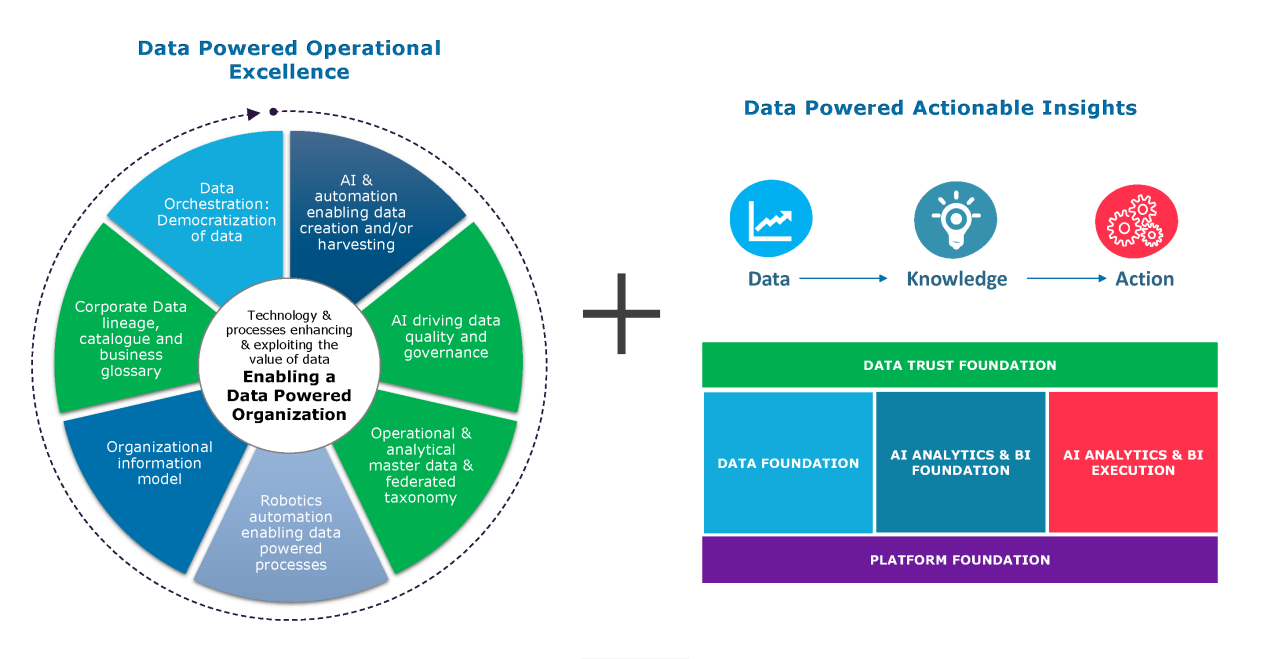
At a time when their strategic insight is badly needed, data specialists still invest too much of their time trying to manually catalogue and manipulate data into central repositories, data warehouses, and data lakes, both on-premises and in the cloud.

To drive trusted analytics and bring accurate insights to the surface, organizations have to start making operational improvements to how data is ingested, catalogued, governed, and shared. Transforming the data fabric starts with the generation and capture of data. It ends with actionable insight.

Time for hands-free data management



The reality is, in the rush to exploit the cloud to reduce operational complexity, maintenance and spiralling costs, new cloud infrastructures haven't made the task of integration and governance any faster or more cost-effective.



Sourced: Capgemini

Data exists in too many places: in inconsistent taxonomies, in huge volumes, across operational and analytical systems.

Data is the oil in the engine of productivity – if it can be properly harnessed. But moving to a data-powered culture is a continuous journey that enables organizations to put the customer first, drive for operational excellence, and find new routes to market. It means finally leaving manual, process-based business cultures behind.

In this paper, we examine four ways to maximize data's value by creating a cloud-native, self-integrating backbone for modern data management.

Hands-free data

The challenge for today's enterprises is to unleash the power of their data. They're moving analytics workloads to the cloud and modernizing applications to boost efficiency and productivity. Data is being democratized and shared with more enterprise users, extending access to analytic insights. These can become the basis for new applications and identifying more opportunities to create customer value.

The last century's IT mantra was to get the right KPI to the right person at the right time. Now its about getting the right data to the right person at the right time – in the right way.

Data needs to be correct and in-line with regulatory, statutory, and organizational governance, without human intervention.

If you consider how much enterprise data can come from sources outside the business – from the vendors of vendors and customers of customers – it's no wonder organizations struggle with discovering, understanding, and trusting the quality of their data. Legacy on-premises systems can hobble things further by imposing inefficient processes. Too many lack the agility to deliver time-sensitive data insights quickly — an absolute requirement for staying competitive.

This rear-view mirror approach to analysis and decision making will not close the productivity gap.

To overcome these barriers, organizations are investing in cloud data warehouses, data lakes and, more recently — data lakehouses, designed to store, update, and retrieve highly structured data for data-driven decision making.

But even the lakehouse model faces challenges. It needs enterprise-scale data integration, data quality, and metadata management to deliver on its promise. Without the capability to govern data by managing discovery, cleansing, integration, protection, and reporting across all environments, lakehouse initiatives are destined to fail.

Data management's three deadly sins

01

Falling into the DIY-IT trap

Implementing effective end-to-end data management can require as many as ten different technology solutions. Bolting all those point products together can put IT teams in an endless state of workarounds and customizations. It delays projects, drains resources, and adds cost. Roadmaps need constant revision, and data quality can be put at risk.

02

Too many manual processes

Relying on manual processes to build a data pipeline can lead to errors. Hand-coded systems are hard to keep current, and rarely meet the scalability requirements of enterprises – throttling innovation and adding cost. If a platform or processing engine is altered or updated, you'll have to re-code. If any changes are made to the underlying technology stack, the code may not be useable at all.

03

Sub-optimal cloud data integration

Ensuring you have robust cloud data integration and ingestion capabilities from all cloud vendors is essential. Your ability to manage cloud data has to extend beyond any single IaaS or PaaS offering, and align with the company's strategy and model for multi-cloud deployment.

As businesses look to move their data to the cloud, hand-coding often comes up as a straightforward way to build the data pipeline. But hand-coding can create bottlenecks. It's also a manual process, and its price can go up as complexity increases.

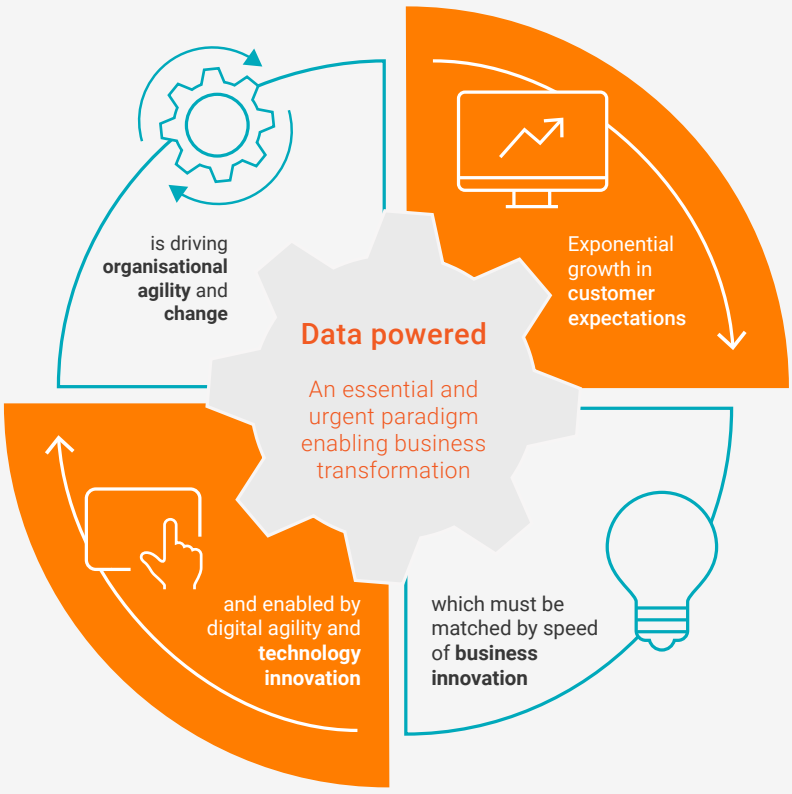
To deliver high-quality, actionable data to the business quickly, you need an AI-driven data management solution that offers a complete view of where all your critical data resides across different silos, cloud repositories, applications, and regions.

Meaningful automation

The stubborn resilience of manual processes is one of the biggest barriers to becoming a data-powered organization. Relying on them limits scalability and creates unnecessary bottlenecks in execution. Manual ingestion and transformation of data, for example, can be a complex multi-step process that creates inconsistent, non-repeatable results.

With scalability and agility at the forefront of business transformation and innovation there is no compeling reason for this to happen anymore. Automated ingestion from known on-premises and cloud data sources is a proven technical approach that adds agility, speed, and repeatability. Automation also suits the fast iteration and flexibility requirements of agile development, because changes can be made very quickly with minimal risk of bugs.

This can become even more vital when data quality is on the line. Analytics platforms should inherit clean and compliant data from the operational systems that feed them. Problems that aren't caught early during ingestion can cause broader downstream issues.



The stubborn resilience of manual processes is one of the biggest barriers to becoming a data-powered organization.

With the growth in data volumes, it is nearly impossible to spot data quality issues on a manual basis. In contrast, using AI to automatically detect signals of incomplete and inconsistent data using automated business rules can have a dramatic impact on the reliability of analytics.

AI needs data, and data needs AI

There are four cornerstones for managing data in the cloud and delivering successful outcomes for big data analytics and AI projects. Applying each one will help avoid the data management pitfalls of the past.

01 Metadata management. First, you need metadata management to effectively discover, classify, and understand how data is proliferating through your organization. Informatica Enterprise Data Catalog (EDC) can help you discover and inventory data assets across your organization. That includes business glossary and lineage data, so you know where data came from and what parts of the business connect to it.

02 Ingestion, curation, transformation and sharing. Next, you need data integration. Data integration is more than simple ingestion; a best-of-breed solution supports all data ingestion and integration patterns. Mass ingestion of files, IoT streaming data, and database initial and incremental loads are vital requirements to hydrate your data lake. Look for ETL/ELT and pushdown optimization to process data once it's in the cloud, ideally performed in a serverless elastic scaling runtime. You also need the broadest connectivity across cloud, SaaS, and on-premises applications.

03 Data quality. Embedding data quality enables you to deliver trusted data through comprehensive profiling, data quality rule generation, dictionaries, and more. Informatica Cloud Data Quality (CDQ) helps you quickly identify, fix, and monitor data quality problems in your cloud and on-premises business applications.

04 Data privacy and security. Lastly, data needs to be protected. When operating in co-located cloud environments, data access and use must be trusted. Applying data-centric protections such as data masking can help limit exposure to appropriate applications and users. This is even more critical in public cloud-hosted application environments, where multiple tenants can coexist on shared resources, to increase risks.

AI requires data to function, but it turns out that data needs AI as well.

With industry-leading metadata capabilities to speed up and automate data management functions, the three-pillar-plus framework can help enterprises avoid past data management mistakes and put themselves on the path to becoming data-powered.

AI and data engineering: a lakehouse with all mod cons

Whatever cloud providers your business uses—Microsoft Azure, Amazon Web Services, Google Cloud Platform, Snowflake, and/or Databricks Delta Lake—understanding all the data assets at your disposal can be challenging.

Centralising everything into a single environment for data consumers to pull from is one approach. But data lives in multiple locations, formats, and cloud environments—and sometimes outside the organisation itself.

Before deciding what to physically bring into the lake, job one is discovering what raw data exists across the enterprise.

AI techniques that automatically scan and index data assets can be used to catalogue all that disparate data. This is a crucial step to extract maximum value from enterprise data assets. Machine-learning techniques can be used to correlate for similarities between different data assets.

After cataloguing is complete, data has to be organised and governed so that business analysts and data scientists can easily find and trust what's there. Metadata plays a crucial role. Without it, the data lake can become a data swamp.

Drivers for changing an organization's relationship with data

Secure, trusted, ethically managed data is powering organizations' ability to innovate and impacting their bottom line. To achieve this, organizations need to have met expectations and challenges coming from three directions:

01

Customer expectations

Customers expect organizations to embrace ethical AI, which in turn requires ethical data management. Ethical data management is part of the foundation upon which customer trust and loyalty are built.

02

Data regulation

Increased global and local data regulation requires mature data privacy, quality and life cycle management; remembering that not all data is governed equally, but does need to be governed appropriately.

03

Data-powered operational excellence

The explosion of data requires agile management and governance while ensuring appropriate levels of data security and privacy. Data is driving operational excellence, new routes to market and innovation without constraint. Increased exploitation of AI and cognitive computing is both demanding and supporting better data quality and data orchestration.

Applying governance and assuring privacy

A machine-learning-based catalogue that provides a clear picture of the data — with a complete view of metadata, summary, and detail lineage, profiling statistics, and 360-degree relationship views — is crucial for leveraging data in the data lake and beyond.

Data governance and privacy policies can also be applied so that you not only know what data you have and how it correlates to other assets, you'll also know what information you're allowed to use and in what ways, as well as who else uses it and how.

To unleash the productivity inherent in data, it also needs to be handled responsibly and in-line with data governance policies for appropriate use. Privacy mandate enforcement has to be part of the ingestion process.

Board-level concerns around digital transformation are often about balancing utility (new products and services, operational efficiency, and value creation) with the risks of abuse or loss, or fitness of purpose quality.

Get it right, and the outcomes are more trusted insights made more widely accessible through data democratization. More workloads get migrated to the cloud, while customer loyalty is strengthened by meeting user demands around fair use.

Conclusion

The average lifespan of an S&P500 company is now less than 20 years[4] – down from 60 years in the 1950s. That calls for a radical shift in strategies and business models if large organisations are going to survive.

Grocery stores are launching banks. E-commerce businesses are building brick and mortar locations. In such a disrupted age, the difference between longevity and irrelevance is being defined by how effectively companies use data.

AI-driven data management keeps pace with accelerating change by unlocking trusted data. It's the crucial step to becoming data-powered, by building a data environment that can deliver trusted, timely, and compliant data to the right people at the right time.

Driven by machine learning and automation, this new model curates all critical data management processes, from data ingestion to data preparation and data governance. It works across multi-cloud environments to accelerate the delivery of trusted data to your business leaders. Stakeholders from across the business can discover ever greater volumes and varieties of data, and turn them into trusted sources of high-value information for more data-driven decision making.

A data-powered transformation is about changing the way organizations, employees and customers think, value and engage with data.

With AI-engineered data management from Informatica and Capgemini, you can finally unleash the power of your cloud data warehouse, data lake, and lakehouse — across on-premises and multi-cloud environments and for data residing inside the enterprise and out.

You can move to become a data-powered organization from a new foundation, resting on four cornerstones that enable you to deliver on the promise of digital transformation.

Contributing author

Fiona Critchley

Global I&D Portfolio Lead Data Foundations & Data Estate Modernisation APAC I&D Portfolio Lead AI & Data Engineering, Capgemini

References

[1] Financial Times, 23 Dec. 2019: Royal Statistical Society's 'Statistic of the Decade'.
[2] Financial Times: https://www.ft.com/content/70cace2c-1bec-4c02-a23c-9d55884b21ac
[3] World Economic Forum: How much data is generated each day?
[4] Credit Suisse: Corporate Longevity: Index Turnover and Corporate Performance

Capgemini

About Capgemini

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations.Capgemini is driven by the conviction that the businessvalue of technology comes from and through people. It is a multicultural company of over 200,000 team members in more than 40 countries. The Group reported 2018 global revenues of EUR 13.2 billion.

Learn more about us at: www.capgemini.com

Informatica

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead - in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer - and unleash the power of data to drive your next intelligent disruption.

Worldwide Headquarters 2100 Seaport Blvd., Redwood City, CA 94063, USA Phone: 650.385.5000, Toll-free in the US: 1.800.653.3871