

Build a Modern Data Team



A new kind of data team

As businesses increasingly move their data and analytics to the cloud, the specialized skills they need to turn data into value are changing at staggering speed. Managed tools that leverage the power of the cloud let us handle more varieties of data, in larger amounts, at greater speed and scale than ever before. But the platform and the tools aren't the only things that have changed – the techniques and resources on our data teams are shifting, too.

In the on-premises world, companies invest heavily in engineers with the expertise to build custom data ingestion workflows, monitor databases, identify data and report on it for business users, and ensure that the data on hand is consistent, up-to-date, and trustworthy. There aren't many alternatives to hand-coding. And the more data and jobs you have, the more engineers you need on hand to do that hand-coding and ensure usable results.

Shifting skills and demands

In today's hybrid and cloud-first environments, though, the highly technical coding skills and other hands-on tasks that were in high demand even a few years ago just to keep workflows going are giving way to new skills and demands. Even low-code and no-code tools and



technologies that were cutting edge three years ago are outdated – designed for an earlier generation of big data, but not one in which the volume, velocity, and variety has continued to get bigger, faster, and more varied. And that means engineers who once had to take occasional classes to brush up their skill sets now need to engage in continual learning just to keep pace with an industry that seems to evolve almost daily.

As technology changes, so should we

We have more data than ever, more things we want to do with it, and more cloud platforms to let us do what was never before possible. But data teams often look the same as they did in the old on-premises years. To make the most of our new data, we also need to rethink the data team – both what it looks like and what it does.

Let's explore the changes necessary to build a modern data team. We'll discuss the challenges your organization will have to address, like reassuring existing team members that they won't be laid off and finding innovative ways to bridge the skills gap. We'll also look at how upskilling and expanding the team can reveal new opportunities to generate more business value from your data.

What's changed?

More data, less structure

Technology market research firm IDC predicts the torrent of semi-structured and unstructured data streaming in real time from IoT sensors and devices will more than quadruple the amount of data in the world in the next five years, reaching 175 zettabytes – or 175 trillion gigabytes – by the end of 2025.¹ An IDG Marketpulse survey we conducted in 2019 found that, on average, data volumes were growing by 60 percent per month at enterprise organizations. (Some companies reported data growth of 100 percent per month.) Companies reported that they used data from 400 different sources.² There simply aren't enough engineers in the world to hand-code all of the integrations necessary to manage that almost unimaginable amount of data. Even if universities were pumping out engineers at breakneck pace, human beings couldn't write code fast enough. It's not humanly or financially feasible to keep doing things the old way.

Cloud data warehouses are here to stay

While companies still need to prepare data for artificial intelligence and machine learning, cloud data warehouses like Amazon Redshift, Google BigQuery, Snowflake, and Azure Synapse Analytics remove the burden of normalizing data to make it fit for use. The line between structured and semi-structured cloud data warehousing and unstructured cloud data lakes has become so blurry that Azure Synapse Analytics has effectively erased it. Business users no longer need to query the right data type, or even know what the right data type is, in order to find the right information to support the analytics they want to perform.

Moving fast in all directions

A data team is also under pressure to perform every stage in a data journey simultaneously. Instead of driving automated decision making and predictive insight by moving smoothly through one advanced cloud analytics project at a time, the data team may be simultaneously ingesting data from social media for one initiative, modeling structured data from back-end systems for another, and working with data consumers to define business rules for a third. Without an understanding of the data itself, they may end up reverse-engineering reports from the desired results and arriving at inaccurate conclusions.



¹ "The Digitization of the World from Edge to Core," IDC, November 2018

² "Optimizing Business Analytics by Transforming Data in the Cloud," Matillion and IDG Research, October 2019

continued...

The rise of the citizen data professional

Business teams no longer have time to write up requirements for a report, share them with IT, and wait until IT has time to pull the data and perform analytics. They want to access and analyze data themselves so they can arrive at conclusions sooner and make strategic decisions faster. Given the global shortage of data scientists and the sheer amount of data available for analysis, it makes sense to decentralize and democratize access to data so business users can take more responsibility for their own data-driven projects. However, the cloud-based tools that enable self-service data access may also be interfering with productivity and the ability to innovate. They often keep companies from seeing and understanding what data these “citizen data professionals” (CDPs) are using and how. Companies need more governance and greater insight into what business users are doing so they can support projects that are innovative and relevant – and cut off resources to projects that aren’t.

Moving skills into the cloud

In short, a data team focused on hand-coding and retrofitting technologies for cloud migration is not necessarily one that is prepared for modernizing data architecture and, with it, cloud data management. The modern data team also needs to include people with the ability to organize data for cost-effective analysis and reporting that leads to intelligent conclusions – that is, data engineers with cloud training and experience as well as the ability to understand the data itself.

Closing the skills gap

Pairing experience with the right skills

It's no secret that universities aren't graduating enough people with the relevant technical skills to meet demand. Just as important, technical graduates who are emerging into the job market inevitably lack domain-specific skills. Someone can graduate with a deep knowledge of data lakes, for example, but no experience applying that knowledge to manufacturing, or medical research, or global shipping. Companies are grappling with a data skills gap fed not just by a lack of tech workers, but by a lack of tech workers who also have industry-relevant context. To bridge this skills gap, organizations are likely to find themselves retraining current employees. A company that migrates to a cloud-first strategy may no longer need on-premises network administration and database engineering skills, but the highly skilled people currently in those positions can learn to manage data lakes and administer cloud services. In retraining those skilled professionals, the company also retains the industry-specific and company-specific experience and knowledge that are hard to teach and harder to replace.

Modern resources for training

To retrain the quality team members they already have, managers will need to determine which learning tools will be most effective in bringing people up to speed. For example, data science still has few formal continuing education classes. But there are several less traditional resources that can help your data team stay current with the latest skills and practices. In other words, the skills aren't all that have changed. The formal ways of acquiring these skills have also given way to a different, more communal mode of training and development.



continued...



kaggle

Managers may want to look at online education on a site like Kaggle, which includes short courses on tools like Python, geospatial analysis, and natural language processing, as well as problems and datasets to solve them. Kaggle also includes community-generated content where people share best practices as they develop them.



stackoverflow

Managers may also want to direct employees to online developer communities like Stack Overflow, where they can ask tightly defined questions, crowdsource best practices, and see which answers have been upvoted by the community as the best.

These learning tools are useful resources for retraining existing staff. But they are also great tools for companies who need to ensure that new hires are up to speed. Even new graduates might not have the most up-to-date skills in this rapidly changing field, while new hires who lack all the necessary data and analytics skills when they first come onboard can quickly make up ground if they have a passion for technology, experience in complex environments, and a proven ability to adopt new tools quickly and often.



In fact, flexibility and adaptability are critical for every member of the modern data team. Teams need data architects with broad on-premises and cloud experience to support a gradual transition to the cloud – or to maximize ROI from existing infrastructure by creating a hybrid data architecture. They also need data architects who are comfortable designing and building with a modular approach. The ability to swap in new technology and capabilities as they emerge and evolve – which may be very often – is vital to future-proofing data ingestion, processing, and analytics.

Does modern mean automated?

By definition, modern data management involves more automation of previously manual processes and less need for human intervention. On the one hand, that's a welcome change, because it means the modern data team will be free to spend less time focused on mundane tasks in favor of innovating, optimizing infrastructure, and thinking strategically about data.

But automation, and reorganization of the data team in general, will inevitably make the people who are currently managing data wonder about the impact on their jobs. They will have questions:

- How can they prepare themselves?
- Will their pay drop when jobs and skills change?
- Will they lose their jobs?

Managers can share this good news with them: People with technical skills who can quickly adapt in a changing environment are likely to remain in high demand for some time to come. A large organization with many IT resources generally has so many silos and pockets of data, and so much data in general, that it will have a lasting need for technically literate people to extract, curate, clean, and load that data into a single source.

Transformation requires a human touch

Though data tools can automate several steps around extracting and loading data, human beings still need to decide what data to extract and where to load it based on the project being supported. This is especially critical when data is being extracted from multiple existing source systems and must be loaded into the most appropriate target system (a cloud data warehouse, data lake, or other) for a given workload.

Data engineers will also need to determine the business logic that helps direct what data you need to join and transform, what analytics to perform, and how that data is visualized. For example, a business may be able to use machine learning to suggest how data from Hubspot might relate to data in Salesforce, but only a human being can create the right key to combine the data so that an analyst can make the best use of it. Also, an engineer can transform that same input data from Hubspot and Salesforce in multiple ways for different outputs and different use cases. The business will always need a technical team that can determine the right transformation for a given business goal.

Technology and business roles converge

Over time, technical and business roles will begin to overlap, which will force members of both teams out of their comfort zones. Business people will need to become data literate and data-driven, and they'll be expected to acquire basic competency with technical tools. With that knowledge, they can do things that they would previously have asked IT to do for them, such as preparing a dataset for analytics. Technical people, for their part, will need to speak fluently about requirements and needs for business projects and be able to communicate with customers in a way that has traditionally been reserved for line of business employees. For team members willing and able to stretch their knowledge and interact more with their IT or business counterparts, this creates great opportunities for augmenting their CVs and driving the business forward.

Some technically oriented people are less comfortable acquiring "soft skills" to complement their IT know-how. Rather than letting them go, a company may simply move them off the data team and assign them instead to operational projects that don't focus on digital transformation. However, if these employees have technical skills the data team requires, it might make sense to partner them with a business team member who can acquire technical understanding that would otherwise be lost while the technical person gradually absorbs the business context.

Anatomy of a modern data team

These are the essential capabilities for the data team of today:

- Core SQL skills for using and managing cloud data warehouses
- Ability and experience with deploying and securing cloud infrastructure
- Familiarity with frameworks for data orchestration
- Strong understanding of applying and documenting business logic through transformations

These capabilities do not necessarily map directly to specific jobs or individuals. However, including them all is likely to result in a data team that includes the following roles:



DATA ENGINEER

The modern data engineer is responsible for keeping data pipelines running and making sure other team members can access the data they need. Data engineers oversee cloud migration and cloud data transformation and may also develop applications for other data team members, as well as data and systems to drive AI and machine learning.



DATA SCIENTIST

Data science is on everyone's list as one of the fastest growing job fields today. Traditionally, data scientists have been centralized in IT, but they are increasingly distributed across teams throughout the business. Data scientists create new ways to work with data and derive value from it, such as data curation or advanced search, matching, and recommendation algorithms.



DATA ANALYST

Data analysts query and report on data in the data lake or data warehouse and use findings to create interactive charts and dashboards for business users to draw on in reporting, diagnostics, and decision making. By turning business decision makers into CDPs, increased self-service data access effectively increases the number of data analysts on the data team.



**CITIZEN DATA
PROFESSIONAL (CDP)**

CDPs may not be a regular fixture at most companies... yet. But the demand for data scientists and the distribution of the data science discipline across various lines of business will increase the number of data-driven and data-savvy employees. CDPs won't supplant a technically trained data team. Rather, they will augment and enhance the way we use data and insights in modern business.



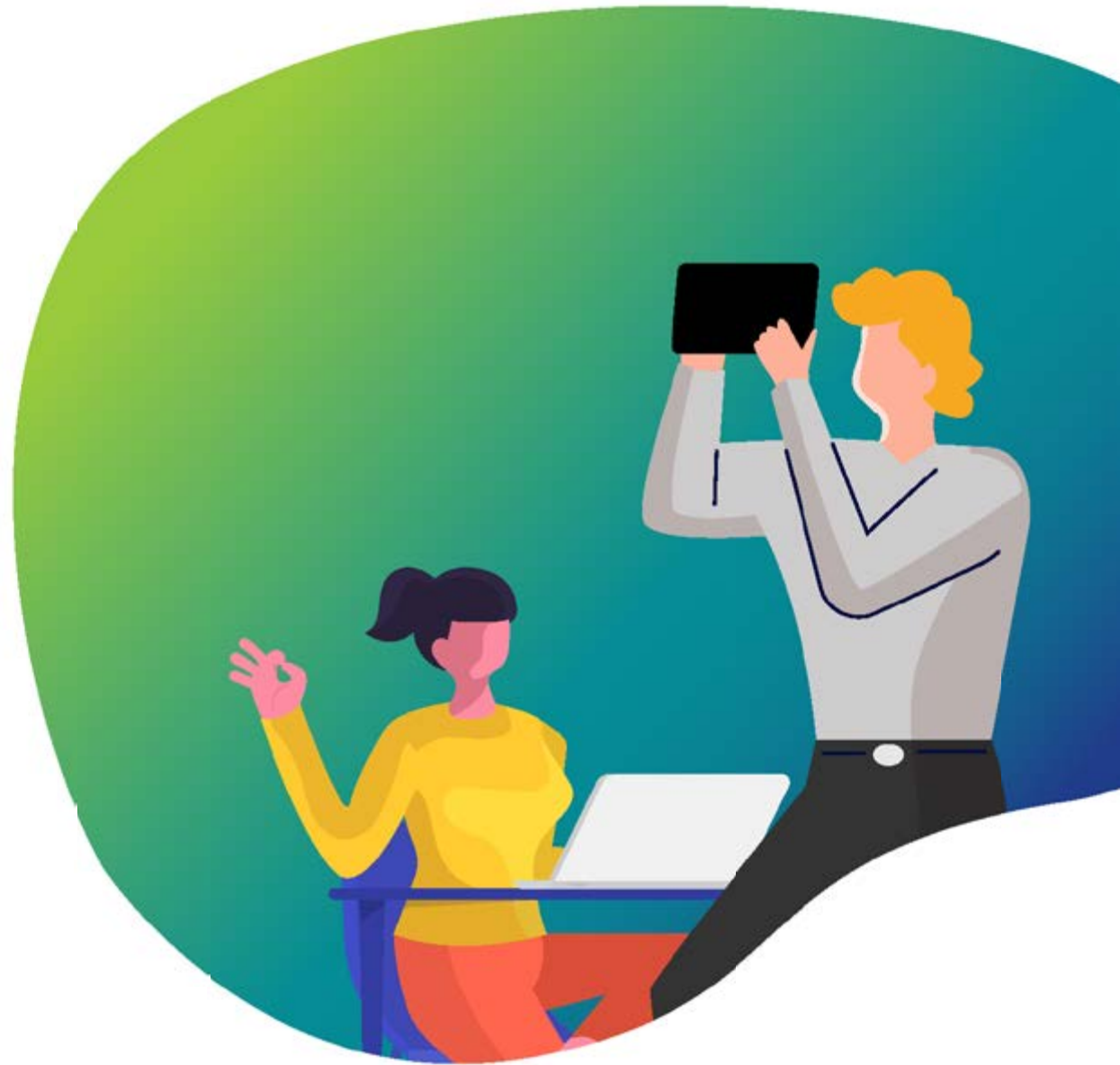
**YOUR EXTENDED
CLOUD DATA TEAM**

Protecting and managing access to data in the cloud is a matter of governance rather than sharing copies of data. This makes it easier to fill the skills gap on your data team with non-employees. Consider leveraging expertise from employees of SaaS partners, consultants who help manage certain types of data or systems, and even solution vendors who lack access to your actual data but provide solutions for managing it.

Is the CDO here to stay?

Although the data team as a whole will have day-to-day responsibility for keeping data flowing, governed, and accessible, someone needs to develop the company-wide data strategy and ensure that every department's initiatives are aligned with it. This role is so important that it's worth being placed in the C-suite, reporting to the Chief Information Officer (CIO) or Chief Technology Officer (CTO). The question is less whether a company should have a Chief Data Officer (CDO), but what the CDO should do.

The CDO role is new and therefore very much a work in progress. Its responsibilities, mandates, and importance vary wildly and change rapidly, both within and between organizations. Yet Gartner predicts that by 2021, the **CDO will be a mission-critical function at 75 percent of large enterprises.**³ These CDOs will not just be in charge of managing data. They will provide insights that support new business models, identify opportunities for innovation, and contribute to revenue. In other words, they will have ultimate responsibility for turning data into business value.



³ "3 Top Takeaways from the Gartner Chief Data Officer Survey," Gartner, January 29, 2019

Governance and the modern data team

Giving business users the ability to select and work with their own datasets doesn't reduce the data team's focus on governance. If anything, it adds another layer: the modern data team may have to govern their own data management processes and data usage, while also governing self-service data access processes and usage for CDPs and other non-IT data consumers. Accordingly, managers redesigning their data team may want to include people and roles that oversee these critical security and governance needs:

- Creating an audit trail to track who does what and who has oversight of the data pipeline
- Establishing data governance processes to ensure that only allowable data is loaded into the data warehouse, data lake, and downstream tools
- Putting appropriate people, policies, and procedures in place to maintain and promote data security and data stewardship



IT needs to guide the CDP

It can be natural to assume that the CDP will eventually replace IT in data management and analytics, but it's more likely that IT will govern and inform the CDP around data, and that won't go away any time soon.

At first, IT may need to build a sandbox for CDPs to learn to use lightweight tools with GUIs to build workflows. IT may also need to provide guardrails and requirements to teach CDPs about necessary concepts like versioning.

As people become more capable of identifying the data they need and retrieving it for themselves, more non-technology managers will be capable of working with data and developing business rules to apply to it. That will free IT to evolve away from narrowly specialized skills and toward a focus on adding new analytics capabilities as quickly and intelligently as possible. But IT will still need to oversee how data is accessed and used in an organization.

The data team in five years

By the time we reach a 175-zettabyte world, data teams will have become increasingly decentralized. Data engineers will need to be less “hands-on” as handling data becomes a part of everyone’s job. Every employee who needs to use data will be data-literate to some extent. Some of them will be CDPs who are comfortable using low-code or no-code applications to integrate data for use in their own complex business intelligence tools, working with IT to mine and explore data.

As data, business, and IT increasingly overlap, the ultimate goal of the data team will be to create a strategic vision around how to use data – and provide the self-service access to data that lets the company achieve that vision.





Matillion and the modern data team

Matillion helps organizations of all sizes build a modern data team by enabling them to achieve new levels of simplicity, speed, scalability, and savings in the cloud.

Matillion products are purpose built for cloud data warehouses, enabling data professionals to achieve faster time to insight and allowing data teams to focus less on hand-coding data pipelines and transformations, and spend more time optimizing infrastructure and helping the business innovate with data.

Matillion Data Loader

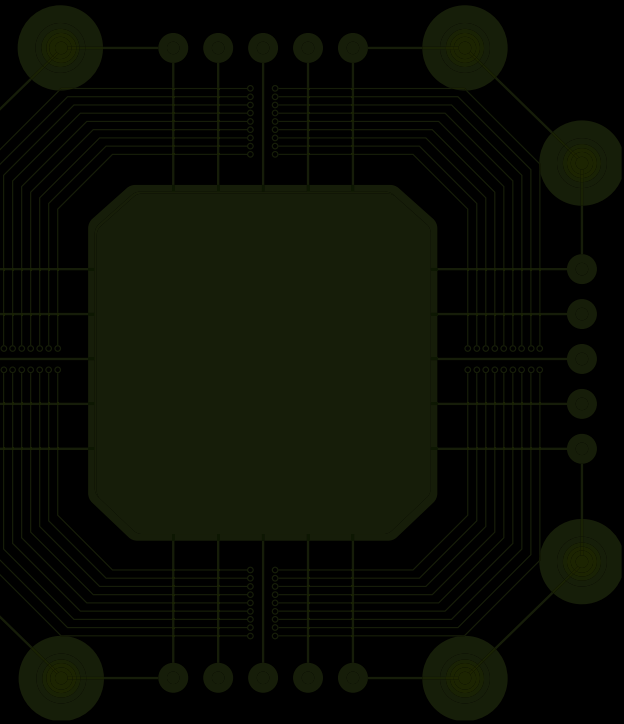
is a code-free, no-cost way for data teams and citizen data professionals to move their data into the cloud.

[Get Started](#)



Matillion ETL enables data professionals to extract data using pre-built connectors, rapidly load that data into the cloud, and transform it using an intuitive, highly visual GUI. Matillion ETL leverages the speed and scale of the cloud, and the native features of Amazon Redshift, Google BigQuery, Azure Synapse Analytics, and Snowflake.

[Request a demo](#)



 MATILLION

