



Modernize
your data warehouse



Table of contents

- 2 Change is your friend
- 3 Why move now?
- 4 Modern cloud data warehouses to the rescue
- 5 Data storage for ALL of your data
- 6 Better workload management
- 7 Quick set-up and time to insight
- 8 How do on-premises and cloud compare?
- 9 Lessons for successful cloud data migration
- 12 Go forth and modernize
- 13 Matillion and modernization

Change is your friend

The Greek philosopher Heraclitus said, “change is the only constant in life.” It’s certainly true of data. Data professionals of all kinds embrace this tenet, developing new architectures and clever patterns, while taking advantage of the latest technologies, to clean, organize, store, and transform data into actionable information.

This is our passion, to find and create new value from the byproducts of doing business and combine them with as yet undiscovered treasure troves. Each new data source is a new opportunity to learn about your products, services, customers, and the market. Your business has endless possibilities hidden within this data - it’s your challenge to find them!

How well you can explore, consolidate, and enrich this data - and how fast you can do it - is what gives your organization a competitive edge. Legacy warehouses were never designed to handle the pressures of modern data demands. They slow down data acquisition and ingestion efforts and struggle to keep pace with your data demands. Growing varieties, volumes, and velocities of data can strain your warehouse and existing BI processes, unearthing platform limitations that are not easily overcome and are expensive to work around. These limitations impact your business’ ability to be proactive and agile and truly leverage valuable new data insights.



It’s not all doom-and-gloom, however. Modern data warehouses, provisioned and procured in the cloud, can solve some of the most difficult challenges inherent in today’s data landscape: poor performance, high costs, increased volumes, and increased variety, all resulting in overall system fragility.

How do you modernize your data warehouse? Start with implementing a cloud data warehouse that can keep pace with your growing data needs, and then explore solutions that are purpose-built to work with them.

The term “legacy warehouse” encompasses on-premises and cloud-deployed transactional databases, like Oracle, MySQL, and Postgres, and data warehouse appliances such as Teradata and Netezza.

Why move now?

If you're on the fence about modernizing your data warehouse, consider this: Data warehouse technology hasn't changed much in almost 40 years, the equivalent of eons in the fast-paced technology sector. During that time, it's performed admirably as the centralized repository for enterprise-wide analytics, capable of storing and processing comparatively large and homogenous data sets. Support for strong-typing, enforced constraints, and CRUD (Create, Read, Update, Delete) operations was essential for its wide adoption and ultimate success. But in recent years, the advent of 'Big Data,' and a voracious appetite for data in general, has exposed some technological gaps that legacy warehouses just aren't equipped to handle.

Massive data volumes require scalability

Organizations are continually looking to new technologies, such as Internet of Things (IoT) devices and log mining, to help answer critical business questions. These sources generate petabytes of data. IDC predicts that the amount of IoT data that is analyzed and used to change business processes in 2025 will reach nearly 80 zettabytes⁽¹⁾. In the near future, legacy warehouses simply won't be able to contain the amount of data produced.

Businesses are already anticipating the data deluge - 71 percent of enterprises expect investments in data and analytics to increase in the next three years and beyond ⁽²⁾. In the past, you'd react to this increased demand by building out existing infrastructure. The velocity of data growth, combined with long implementation times and exorbitant costs, render this scalability strategy unsustainable. The static nature of bare metal resources that house legacy warehouses creates a constant tug-of-war between under and over utilization, ultimately leading to poor performance or wasted dollars.

The Hadoop ecosystem was developed to solve some of these problems, but it comes with its own challenges. Adoption demands a large investment in human capital, highly skilled engineers, and technicians to build, scale, and maintain a complex environment. The hodge-podge assemblance of bolt-on utilities, each with their own varied levels of support and community interest, creates a maintenance nightmare with no simple way to gauge the larger platform's efficacy.

Some formats compromise data quality

Web-friendly formats, like JSON, enable applications to pass around complex data structures with relative ease. NoSQL databases were built to store these formats natively, unburdening developers from the strictures of things like predefined schemas, foreign keys, and enforced constraints. This freedom, however, comes with a significant cost to data consistency and cleanliness, confusing even simple relationships that analytic models depend on to be valuable.

In other words, each of these technologies only solves a specific problem, and at the cost of functionality that has made the data warehouse so successful. Until now.

(1) 152,000 Smart Devices Every Minute in 2025: IDC Outlines the Future of Smart Things," Forbes, March 3, 2016

(2) Global State of Enterprise Analytics, 2018," Forbes, August 8, 2018

Modern cloud data warehouses to the rescue

The cloud data warehouse is a rare and welcome innovation in data storage and analytics. Built and delivered as a managed service, cloud data warehouses have been specifically designed to address some of today's most challenging data problems. There are several ways in which cloud data warehouses solve legacy warehouse challenges for businesses of all sizes and industries.

Usage-based pricing models

Historically, legacy warehouses burdened your IT department with the need to laboriously assess and extrapolate your data needs one, three, or five years down the line. In reality, no business can exactly forecast and predict future data needs in order to scale accordingly. You end up in a 'Goldilocks Dilemma': Paying up front for dedicated hardware and software that might not be fully utilized for years; or underestimating, sending you back to your CTO within the first six months to go through the painful procurement process all over again. Even if you can forecast your data needs and build the infrastructure to match scale, it's a finite solution.

In contrast, a cloud data warehouse can be right-sized on a moment-by-moment basis, ensuring that your costs match your requirements now and in the future, procured and provisioned in a matter of minutes.



The ability to shift existing workloads to a cloud data warehouse in a true Proof of Concept, while incurring little or no cost, is game changing. Imagine being able to take your largest datasets and your gnarliest transformations and test them out in a matter of hours, having committed to nothing.

Data storage for ALL of your data

The cumbersome process of monitoring disk utilization and purchasing, installing, and configuring new disks on expensive network attached storage (NAS) appliances is a thing of the past. The cloud data warehouse works with incredibly large volumes in a variety of formats and storage options, solving the problems that arise from increased data demands and complexity, all while simplifying the management of your storage layer. Snowflake has taken a unique approach to this problem by natively separating **storage and compute**, so that each can be managed independent of the other.

Amazon Redshift Spectrum and **External Data Sources for BigQuery** add similar functionality to your cloud data warehouse, as an optional feature, giving you the flexibility to mix and match internal and external storage as you see fit.



Better workload management

The ability to tailor the warehouse to your data analytics needs, on demand, is perhaps its greatest benefit over legacy warehouses. Cloud data warehouses, and their capacity for Massively Parallel Processing (MPP), are built to handle intense workloads that can be scaled quickly and easily. With Snowflake, you can **resize** your compute resources (called a **Virtual Warehouse**) with one command. With Amazon Redshift, you can **scale elastically**, adding or removing nodes in a cluster using a simple API call, or their easy-to-use management console. Google BigQuery takes advantage of **features** like serverless technology and auto-scaling, effectively reducing workload management efforts to nothing.

Matillion is purpose-built data transformation for cloud data warehouses. Whether you are ingesting data, transforming it, or writing it out to cloud storage, Matillion optimizes your workload management with native functionality to improve performance (and hopefully save you time and money on your queries).



Quick set-up and time to insight

On the outside, the cloud data warehouse looks like any other. Each comes with its own graphical user interface for quick access, testing, and familiarization. They speak the same language, for the most part, and work with the same tools, so the learning curve is minimal. Under the hood, they don't require complex configuration and infrastructure like other 'Big Data' technologies. Your teams can create proofs-of-concept on all of the major cloud data warehouse technologies, **Google BigQuery**, **Snowflake**, and **Amazon Redshift**, without expensive training or waiting for long procurement and deployment processes.

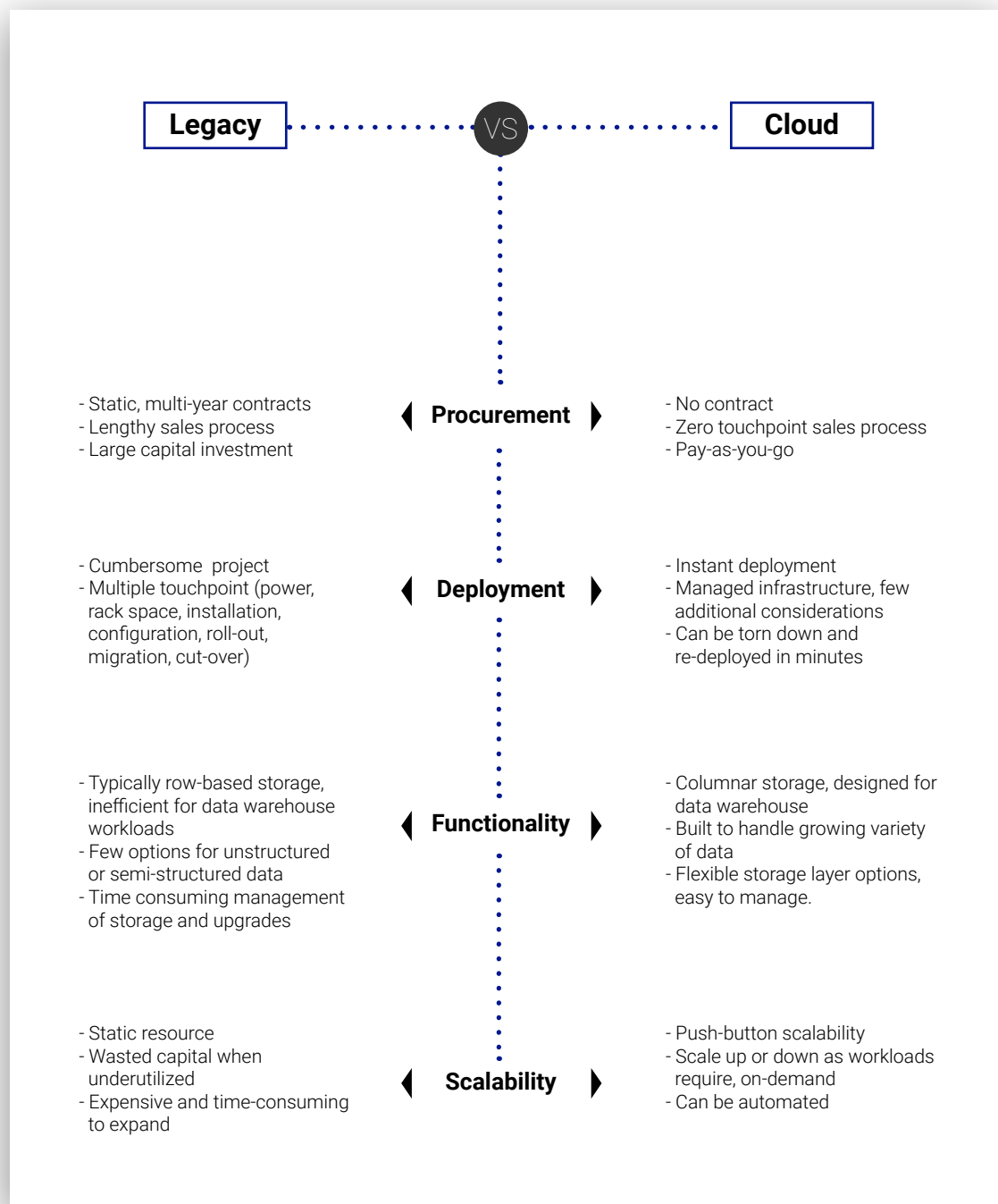
Your people are your greatest asset. They give 110 percent to their craft because they know how important it is, because they love seeing the final product come to life, and because they love the challenge. Unfortunately, they spend large amounts of time and effort chasing minor performance improvements and waiting for other workloads to finish. Imagine what they could accomplish if the limitations of a legacy warehouse were suddenly lifted, and they had a solution as powerful as they are to tackle today's immense data challenges.

No structure? No problem

Integrating existing operational data with semi-structured and unstructured 'Big Data' can be a major technical challenge. Legacy warehouses aren't designed to consume these types of data naturally, leaving you with few good options. Data lakes help solve some of these challenges, specifically related to storage and discovery, but it's still a significant pain to combine, aggregate, and enrich structured and unstructured data with existing operational data.

The cloud data warehouse treats these unique structures as first-class citizens, providing features and functionality designed specifically for the consumption of raw semi-structured data, along with the tools to unravel it. Snowflake has **specific semi-structured data types** so there's no need to shoehorn this data into columns that almost accommodate them. With Amazon Redshift, you can **define external tables** so that the platform "understands" how to represent your data in columnar format, making it easy to combine with other operational or dimensional data. Google BigQuery also supports **loading and flattening** JSON data, with options for auto schema detection to simplify the process even further. In short, these platforms are well equipped to handle the expanding variety of data that organizations are attempting to consume, as well as the ballooning amount.

How do on-premises and cloud compare?



Lessons for successful cloud data migration

Cloud-based models inherently reduce risk. Pay-per-usage models and improved storage and compute capabilities mean that you can spin up, trial, and run an economical proof of concept, largely unmanageable with on-premise resources. It's a fail fast model, and can yield a quick ROI.

Is it too good to be true? Even the most optimistic among us might be thinking, "Surely there are perils and pitfalls waiting around every corner." Not necessarily, but there are five critical things to consider that can make the difference between success and failure. In planning your migration, here are some common lessons we've learned from our customers. Keep them in mind, and there's an excellent chance you'll be in the success column.

01

Prepare for new technical best practices

Not setting expectations can put your initiative at risk. While standing up a new cloud data warehouse is fast and easy, the underlying technology is different from legacy warehouses. Such differences can be overlooked because cloud data warehouses use standard SQL dialects and similar connection options. A cloud data warehouse may warrant new design choices, to take advantage of **columnar** storage and bulk data loading, for example. The subtle differences will impact how table structures are designed, loaded, and queried, requiring a paradigm shift and a small learning curve.

To learn more about the differences between on-premises and cloud data warehouses, check out the Matillion eBook **[From ETL to ELT](#)** which walks through the changes in data warehousing and how Matillion changed their data transformation method to operate efficiently in the cloud.

02

Evaluate all of your options

Leverage the consumption model and the wide variety of free or low-cost trials to evaluate the available cloud data warehouses and complementary solutions you will need to be successful, such as ETL and BI tools. For example, traditional ETL tools aren't built for the cloud. So, while they may technically work with a variety of legacy warehouses, they're unlikely to take advantage of the native improvements and best practices that your chosen cloud data warehouse offers. In fact, it's more likely that they treat your cloud warehouse like your same old warehouse, which can result in some of the same old performance bottlenecks, appropriately causing you to question why you're migrating in the first place.

Instead, evaluate new, cloud-native solutions and techniques, so that when it's time to make the move, you've selected the very best that's on offer for the cloud and data warehouse that suits you. Test things in a variety of use cases, ranging from simple tasks to your most challenging or poorest performing workloads. One of the key benefits to the cloud is the ability to stand up new resources on demand. Embrace it!

03

Bring in stakeholders early on

When evaluating cloud data solutions and tools, include a wide variety of stakeholders if possible, including, but not limited to:

- > Business analysts
- > Data scientists
- > Project managers
- > Heavy consumers
- > Power users

Remember, change is constant. But it's also scary for some. It can be alarming, from an analyst's perspective, to hear that the data they are familiar with will soon become stale and they'll have to develop a trust for a new warehouse all over again. Allowing them to see some of the benefits early on, and to provide their input on existing pain points, could help consumers become champions for the new warehouse, rather than alarmists.

Matillion is purpose-built for each cloud data warehouse its supports, meaning better performance! Ask us how to test out Matillion with your preferred CDW.

04

Create a plan for the Cloud

Amazing news: if you're moving to the cloud, you aren't working under the typical licensing and hardware constraints that would force you to lift and shift your existing warehouse, and all associated processes, all in one go and on a tight schedule. So don't rush--make a thoroughly considered plan.

You can migrate workloads gradually, starting with some of your simple workloads as a way to familiarize yourself with new features and to establish benchmarks. Or, start with the migration of raw data from your existing warehouse, and any other data silos that you may want to centralize. Take the opportunity to explore new design patterns, to clean up technical debt, or both.

Don't just have a plan, have an exit plan. In times past, there was no exit plan because the contracts were signed and the ink long dry before the hardware even arrived at your data center. If things aren't going as expected, you've allowed yourself some room for lessons learned. It should be okay, even encouraged, to change course whenever you need. Go back to your evaluation, pinpoint the missing criteria, and formulate a new plan. The cloud affords flexibility. Take advantage of it.

05

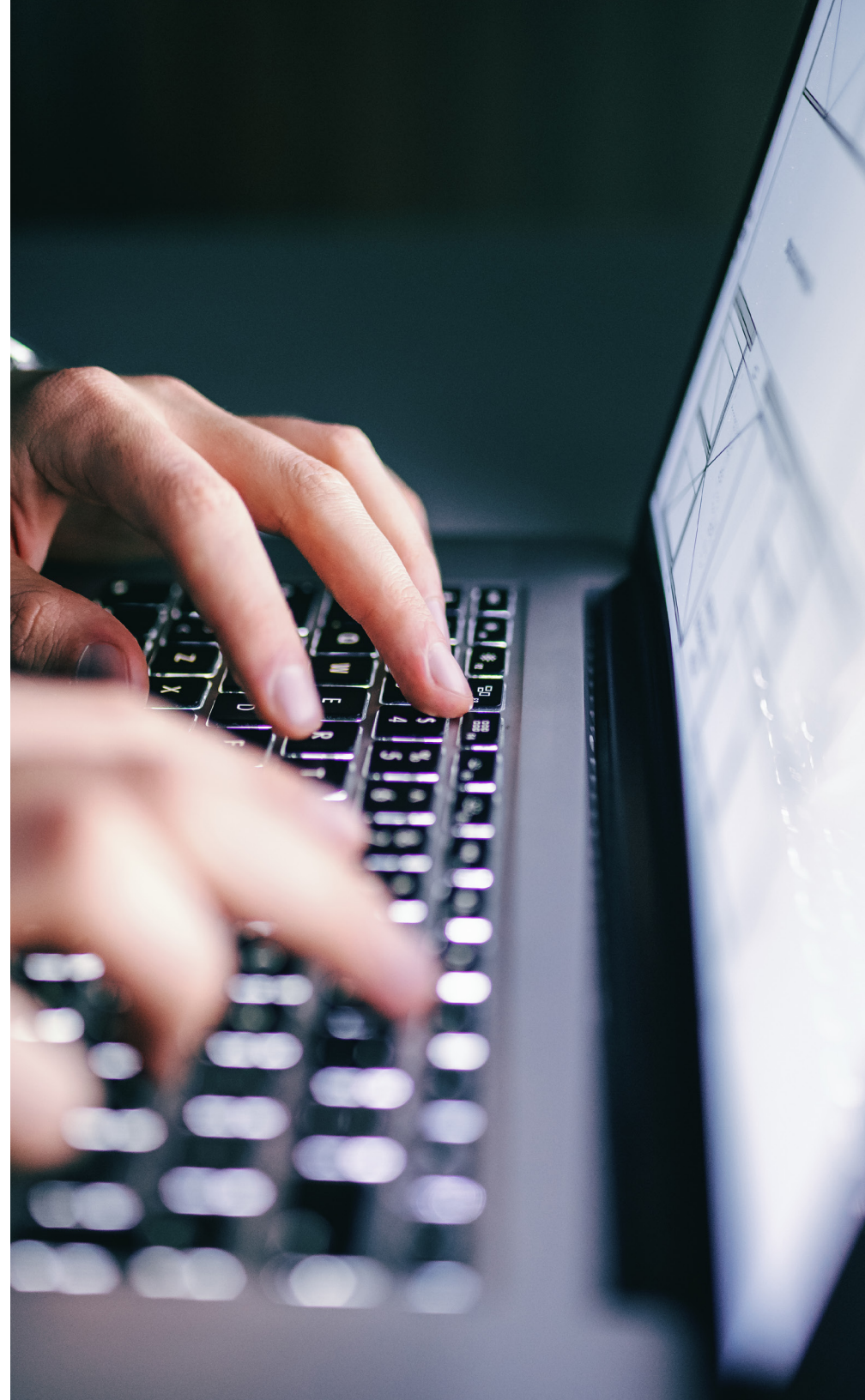
Engage with us

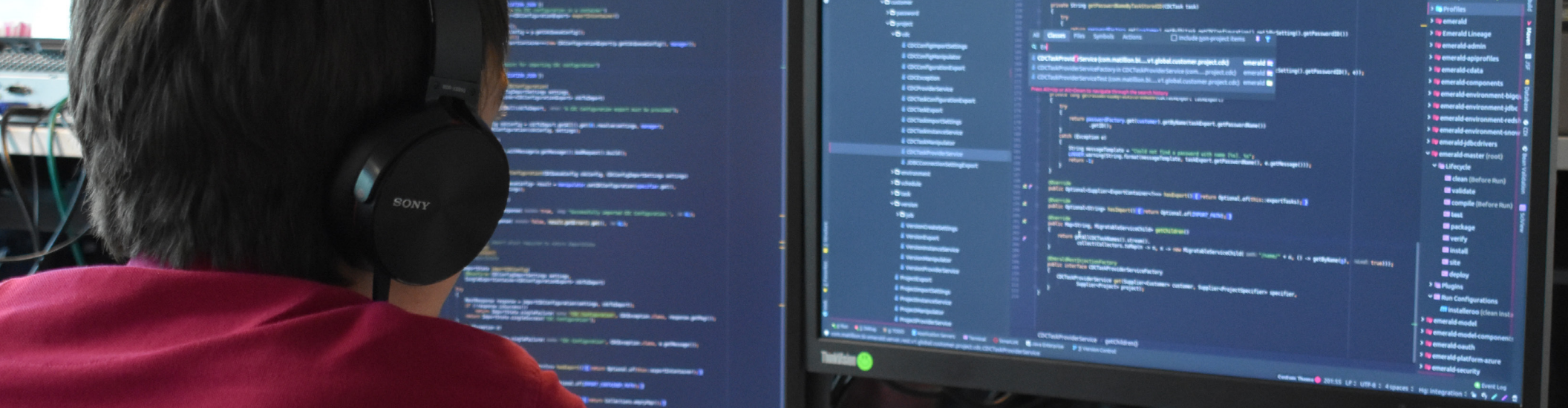
Even with a rock-solid plan, sometimes there just isn't time to figure everything out, despite best efforts. Some teams don't have any bandwidth to spare, so sparing a resource or two for even a few hours a week to work on a modernization initiative isn't an option. We perform data warehouse migrations and we're very good at it. We stay up to date on the latest tooling and are uniquely aware of a variety of pertinent topics that may impact your use cases. And once your migration is complete, we can provide documentation, training, and a proper hand-off to your existing team.

As a Matillion partner, we have the skills and experience to help you with cloud data warehouse migration projects. Contact us today to learn more.

Go forth and modernize

The cloud is redefining how we think about storing and processing data. The power afforded by a modern data warehouse and technology stack can give your business a competitive advantage like never before. You've learned several things to keep in mind as you try out solutions, move to the cloud, and start your cloud-based approach to data analytics. Time to push your business forward and unlock the potential of your data. Change the way you think about your data and technology landscape. By leveraging the power of the cloud, you and your business can explore your data in a new light to unearth insights about your customers, find efficiency in your logistics, and answer your biggest questions better than ever before.





Matillion and modernization

Matillion provides industry-leading data transformation products for cloud data warehouses. Delivering a true end-to-end data transformation (not just data preparation or movement of data 'as-is' from one location to another), Matillion provides an instant-on experience to get you up and running in just a few clicks, a pay-as-you-go billing model to cut out lengthy procurement processes, and an intuitive user interface to minimize technical pain and speed up time to results. Matillion is available globally for Snowflake, Amazon Redshift, Google BigQuery, and Azure Synapse Analytics on leading cloud infrastructures.

When you're ready to modernize, Matillion is purpose-built data transformation for the cloud that can help you extract and load data from existing systems, in order to conduct a wide variety of transformations on that data, all using the power of the cloud. Procured and deployed directly into your cloud infrastructure, and billed hourly through the marketplace of your choosing, Matillion is aligned with your cloud strategy. Further, Matillion natively employs best practices ascribed by each cloud data warehouse vendor, taking full advantage of platform specific functionality, allowing your team to focus on what they do best.