



How to Enable High-Performance ELT Processing

Automate Your Data Transformation



Big data alone does not guarantee better business decisions. Often that data needs to be moved and transformed so Insight Platforms can discern useful business intelligence. To deliver those results faster than traditional Extract, Transform, and Load (ETL) technologies, use Matillion ETL for Amazon Redshift. This cloud-native ETL/ELT offering, built specifically for Amazon Redshift, simplifies the process of loading and transforming data and can help reduce your development time.

This white paper will focus on approaches that can help you maximize your investment in Amazon Redshift. Learn how the scalable, cloud-native architecture and fast, secure integrations can benefit your organization, and discover ways this cost-effective solution is designed with cloud computing in mind. In addition, we will explore how Matillion ETL and Amazon Redshift make it possible for you to automate data transformation directly in the data warehouse to deliver analytics and business intelligence (BI) insights that meet your business needs.



Introduction

According to the Forrester Report, “Insight Platforms Connect Data to Action¹,” in many cases the way data lakes are being deployed and used will not meet the evolving needs of your business. Although it is true that data lakes enable you to store large amounts of structured and unstructured data, they are still expensive, slow to change, and often contain data of questionable overall value.

While Amazon Web Services (AWS) has helped to transform IT infrastructure into a scalable, cost-effective, and on-demand service, Amazon Redshift has had a similar effect on data warehousing and big data analytics. To harness the power of big data, your organization could benefit from understanding how to configure Amazon Redshift, so it delivers the right balance of performance and cost for your specific use cases. Matillion ETL for Amazon Redshift enables the ingestion of relational and non-relational data from both your existing data warehouse and new data sources. Matillion ETL loads dozens of data types into Amazon Redshift, using more than 50 pre-built source connectors, and transforms the data into an analytics-ready state. AWS makes it possible for you to build a data lake within Amazon Simple Storage Service (Amazon S3), and then query that data for valuable insights.

As your organization continues to generate ever-increasing amounts of data, it is crucial that you are able to maximize value from this information - enabling easy data-driven decisions.

Consider an ELT approach to cut your development time in half and trim months off your projects—and get started with just a few clicks. Case in point, **learn how Matillion has helped customers:**

- Load millions of rows of data using the Massively Parallel Processing built into the Amazon Redshift platform - for some users, this happens almost instantly.
- Transform data from all your disparate data sources —code free
- Model your data for highly performant BI / data visualization
- Rapidly export your relational and Software-as-a-Service (SaaS) data to a data warehouse on AWS for storage and future use
- Use our modern, browser-based environment for full-featured graphical job development

Read on to gain additional insights on how Matillion ETL for Amazon Redshift can help you overcome challenges by simplifying data warehousing and big data analytics workflows, focus collaboration among different teams within your workforce, and capitalize on faster performance and a modern approach to big data.

Challenges

If your organization is using legacy data warehouses to manage large datasets, you may be encountering performance issues when accessing and analyzing your business and customer information. This may be because your legacy data warehouse is unable to handle useful volumes of data and is delivering inadequate insights. It may also be caused by the fact that legacy data warehouses perform poorly when too much data is loaded into them. Although your organization may understand the value that lies in the data itself, data management and integration can be complex, slow, expensive, and often the first major obstacle to a large-scale data warehouse project.

In many cases, it has become apparent that enterprise data centers lack the flexibility and scalability of cloud environments and are often incompatible with cloud infrastructure. This can make the collection, storage, and analysis of data ineffective, and they are often more expensive to maintain than a cloud solution. There can also be difficulty loading different types of data, be it structured or unstructured, into the same data warehouse. Matillion ETL enables easy ingestion of relational and non-relational data from multiple sources into Amazon Redshift.

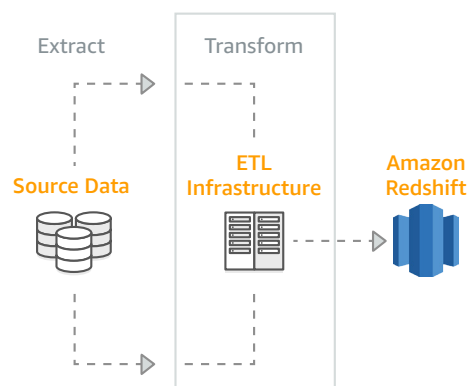
This issue is often illustrated by organizations and solution providers missing opportunities with data lakes, because they are not using the proper tools and solutions to gather and analyze the correct data. As a result, they are unable to evaluate the data they have in a meaningful way to get the results and benefits they are seeking. Matillion ETL for Amazon Redshift helps to overcome this by combining multiple sources of data to extract valuable information and gain important insights.

Matillion ETL for Amazon Redshift was created to help your organization overcome the challenges mentioned while delivering faster results and reducing development time.

ETL vs ELT

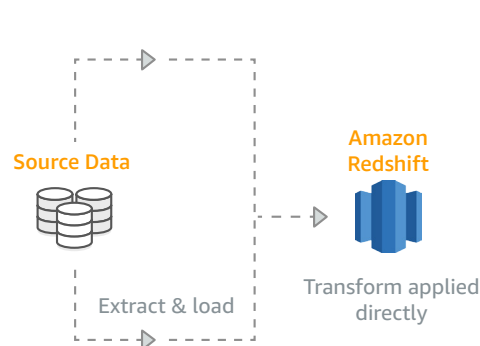
Traditional

Extract, transform & load approach



Matillion

Extract, load & transform approach



Transform your Workflow

Adopting AWS for data warehousing, and moving your data with Matillion ETL for Amazon Redshift, can help your organization simplify workflows with a user-friendly browser-based tool that is optimized for a cutting-edge user experience. Matillion ETL does this by supporting any size/type of Amazon Redshift cluster and incorporating transformation components that enable you to more rapidly execute automated data transformations—including filter, join, aggregate, rank, and table input/output—in a graphical user interface.

The key component of transforming your workflow, however, is found in Matillion's unique cloud-based push-down ELT architecture. By changing the traditional Extract, Transform, Load approach to an Extract-Load-Transform (ELT) model, Matillion ETL for Amazon Redshift accelerates data warehousing processes by delivering results faster.



Simplify workflows with a user-friendly browser-based tool

In short, because ELT has been created with cloud computing in mind, it is more performant and efficient than traditional, time-consuming ETL technologies. By reordering this process, the ELT solutions employed by Matillion ETL for Amazon Redshift enable you to orchestrate data transformation directly on the data warehouse. This can help your organization leverage the full power of AWS data warehousing capabilities to deliver the business insights you need to grow your business.

Additionally, the tool helps you realize the full benefits of your organization's investment in big data analytics. Traditional ETL solutions commonly require excessive compute, storage, and memory resources to execute their transformations. Matillion ETL for Amazon Redshift allows you to put your resources where they'll have the greatest impact, namely inside your Amazon Redshift cluster. While Matillion ETL does require some computing resources in the form of an Amazon Elastic Compute Cloud (Amazon EC2) instance, these resources are minimal and are only used to generate the push-down transformations within Amazon Redshift, where the actual work happens in a more efficient and cost-effective manner.

Reclaim your Time

Amazon Redshift is a fast, scalable data warehouse that simplifies the analysis of all your data across your data warehouse and data lake. It combines machine learning, massively parallel query execution, and columnar storage to run queries across petabytes of data in your data warehouse.

Your organization benefits from Amazon Redshift's faster performance and higher throughput, which means you spend less time waiting and more time analyzing data. Amazon Redshift also scales quickly to meet your needs and is able to quickly analyze any size of data stored in Amazon Simple Storage Service (Amazon S3). Plus you can also resize an Amazon Redshift cluster with just a few clicks, customizing the amount of resources available to your particular computing needs.

Then with Matillion ETL for Amazon Redshift, you can prepare data to be ingested into your BI/visualization tool or insight platform for reporting, dashboard views and interactive analysis.

Insight Platforms Unify Technologies¹

Matillion ETL for Amazon Redshift enables insight platforms to consume data in the right formats for effective analysis rapidly. By optimizing the combined capabilities of each service, it helps you more effectively manage and analyze data, then test and implement the insights that are discovered so they can become business actions.

As your needs grow and your tools evolve to meet them, Matillion ETL for Amazon Redshift delivers a solution that offers big data management, analytics, insight application development, and outcome monitoring and measurement in a single offering. In short, this solution does the heavy lifting to leverage the capabilities of Amazon Redshift to reduce the pressures on the insight tools your organization may be using. Organizations that have deployed this solution report that it has helped them alleviate bottlenecks in their BI tools and processes and reduce costs.

This solution is built for the cloud and enables you to manage and integrate data across many frameworks. Combining analytics and data management tools empowers insight platforms to host custom business logic that automates these processes, to help accelerate the deployment and execution of predictive models.

¹Insight Platforms Connect Data to Action," Forrester; Brian Hopkins, January 2018

Benefits of ETL Adoption

Cloud-native and built to take advantage of the data warehousing capabilities offered by AWS, Matillion ETL for Amazon Redshift delivers results faster than traditional ETL technologies. Reduce your development time after getting started with just a few clicks.

Matillion ETL for Amazon Redshift formats data in many different ways to set up schemas so different datasets can serve multiple teams in your organization (for example, HR, accounting, Dev/Test, etc.). Throughout the project or use case, it is purpose-built to leverage modern data warehousing capabilities. Matillion ETL enables faster and more effective data flow development, which can help your organization save money and get more done.



Enable your teams to work more closely

Make Loading Easy

Using Matillion ETL for Amazon Redshift simplifies the process of loading data into Amazon Redshift, automating many of the steps to accelerate the process. This solution is easily integrated with native AWS services such as Amazon S3 and Amazon Relational Database Service (Amazon RDS). For example, the Amazon S3 load component in Matillion ETL for Amazon Redshift provides drag-and-drop functionality for stored data loads, while the Amazon RDS Query component in Matillion ETL for Amazon Redshift loads Amazon RDS data to and from Amazon Redshift.

Take Control of Your Data

The ease of use—you can build your own ETL processes—delivered by Matillion ETL helps you rapidly transform your data using the power of Amazon Redshift, enabling your organization to quickly gain valuable business intelligence to help shape future decisions. The visual orchestration of sophisticated ETL processes with transactions and loops accelerates the process of gleaning valuable information from the data being processed. This feature is enhanced by the ability to schedule, review, and change your orchestrations to fit your needs.

The integration with other AWS services such as Amazon Simple Query Service (Amazon SQS), Amazon Simple Notification Services (Amazon SNS), and Python increases the flexibility of Matillion ETL for Amazon Redshift by providing more options on how to access, view, and analyze your data.

Facilitate Focused Collaboration

Enable your teams to collaborate in real time by using the live collaboration feature that is part of Matillion ETL for Amazon Redshift. This makes it possible for remote teams to modify the same ELT job in real-time and see each other's changes as they're made.

Matillion ETL's job versioning and import/export features allow large teams to work together, and also make it easier for your organization to apply DevOps practices such as continuous integration and continuous delivery. This helps provide multi-environment and multi-schema support for the control of development, test, and production environments.

Secure Your World

The Key Management Service (KMS) is a service provided by AWS and leveraged by Matillion to help you establish and maintain security best practices. The dependability of KMS, combined with the end-to-end encryption, network isolation, and audit and compliance of Amazon Redshift helps to continuously protect your data, anywhere in the world.

The simplified AMI deployment of Matillion ETL for Amazon Redshift within your Amazon EC2 instance means that data never has to leave your environment. This makes it a suitable solution for data-sensitive industries such as healthcare or finance, since the data is not passed to an ETL/ELT server.

Explore Today

Now that you have learned how Matillion ETL for Amazon Redshift can benefit your organization, see it in action:

Schedule a demo: See how fast and easy it is to Extract, Load, and Transform data with Matillion ETL for Amazon Redshift. Experience how the modern, browser-based interface launches the AMI within minutes, and enables you to load nearly unlimited rows of data in a matter of seconds.

Start a 14-day free-trial: Launch from AWS Marketplace and begin getting hands-on experience and insight into the easy onboarding process, as well as how Matillion ETL for Amazon Redshift can reduce your ETL development time. This free trial includes full support, as well as access to the monitored support portal, help documentation, and video tutorials.

Conclusion

If your existing infrastructure is aging or unable to quickly ingest, process, and/or analyze your ever-increasing volumes and varieties of data, you may be missing valuable business insights in your analyses. Matillion ETL for Amazon Redshift not only works faster than traditional solutions, it also delivers more impactful results. Consider modernizing your data warehousing capabilities by adopting this scalable, secure, cost-effective solution so you can spend less resource on time-consuming and labor-intensive tasks, and instead focus on data, data analytics, and power business insights.

About Matillion

Matillion's cloud-native ELT solutions give fast-growing, tech-savvy companies the speed and power they need to gain critical insights from their data, in record time. Matillion builds purpose-built ELT products for the cloud data warehousing platforms they support, including Amazon Redshift. Recognizing that these platforms offer users unparalleled processing power in their own right, Matillion has created tools to take full advantage of their innate speed and performance - to transform data within the cloud itself.

About AWS

For over 12 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers over 125 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, virtual and augmented reality (VR and AR), media, and application development, deployment, and management from 55 Availability Zones (AZs) within 18 geographic regions and one Local Region around the world, spanning the U.S., Australia, Brazil, Canada, China, France, Germany, India, Ireland, Japan, Korea, Singapore, and the UK. AWS services are trusted by millions of active customers around the world—including the fastest-growing startups, largest enterprises, and leading government agencies—to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit <https://aws.amazon.com>.