

**EPLEXITY**

**aws** partner  
network

# Data warehouse modernization: From monolith to modular

# Contents

<b>01</b>	Introduction
<b>02</b>	Why data warehouse modernization?
<b>04</b>	The characteristics of a modern data warehouse
<b>07</b>	Why data warehouse modernization with AWS and EPLEXITY?

# Introduction



For years, data warehouses have enabled businesses to uncover insights from and make sense of their data by pulling that data from many different sources for reporting and analysis. Traditional data warehouses have monolithic architecture and on-premises infrastructure that are difficult to scale and maintain, offer little flexibility, and incur high costs. With agility now key to almost all aspects of business and IT, this model is not sustainable or cost-effective.

Data warehouse modernization moves the data warehouse to the cloud, allowing it to break free from the many limitations of on-premise architectures. With new design tenets, including a modular architecture, the modern data warehouse exploits the agility, security, scalability, and economics of the cloud. As a result, organizations can get exactly what they need when they need it—fast and at a fraction of the cost of legacy data warehouses. This eBook explains why data warehouse modernization is needed, dives into the characteristics and benefits of the modern data warehouse, and provides next steps for organizations who want to modernize.

# Why data warehouse modernization?



Data warehouses have traditionally resided in on-premises data centers. Their architecture and infrastructure enable business intelligence and analytics on structured data. However, the nature of the data that businesses are collecting has been changing rapidly for several years now. Streaming data, semi-structured data, text, and voice—these and others are all outside the highly structured transactional data box. They can even outweigh the box in terms of offering valuable insights that drive accurate and timely decision-making. But they must be accessed and analyzed properly.

Traditional data warehouses are not designed to support all the new forms of data efficiently or cost-effectively. Their infrastructure is rigid, and the long-running ETL processes they require affect compute and performance. Cloud-based services offer a solution to enterprises who want to break free from the lack of flexibility, speed, and economies of scale of their traditional data warehouses and run analytics on new data types at the velocity needed by today's businesses.

Cloud technology, with its high performance, simple deployment, near infinite scaling, and easy administration at a fraction of the cost of on-premises solutions has opened the door to a new and modern data warehouse construct. This construct removes the limits and restrictions inherent in traditional enterprise data warehouses by supporting rapid data growth and interactive analytics over a variety of data types using a single interface—on the cloud. It automates numerous tasks and its architecture is modular.

Specific characteristics, such as separation of compute and storage, enable modern data warehouses to address the flaws inherent in their traditional predecessors and deliver benefits more suited to modern data types, today's sophisticated analytics tools, and computing workloads.

# The characteristics of a modern data warehouse



In contrast with an expensive on-premises, one-size-is-all-you-get solution, the following characteristics of the modern data warehouse enable organizations to implement a solution that suits their needs, delivering insights only when they need them.



## Separation of compute and storage

In a modern data warehouse, storage and compute are decoupled. Compute resources are deployed on demand when data is queried directly—and automatically when a large number of concurrent queries are run on data loaded into the data warehouse. Meanwhile, core data resides separately so that only the data that is needed is accessed. This is a significant divergence from the on-premises data warehouse whereby any additional compute is tied to additional storage.

With the separation of compute and storage, the modern data warehouse spins up clusters as needed; there is no need to run processes indefinitely. You can automatically add and remove capacity to ensure consistently fast performance, even with thousands of concurrent queries and users. In addition, a future-ready engine enables users to answer complex analytical questions. No more on-premises “stacks and racks and racks.”



## High performance at petabyte scale

Data warehouse modernization offers higher performance without incurring wait-time, resource, or downtime costs. It enables faster data analysis, spinning up a cluster in just minutes. A modern data warehouse will implement techniques such as columnar storage, data compression, and zone maps to deliver data at scale more efficiently. For example, with columnar storage, each data block can hold column field values for as many as three times the records as row-based storage. This reduces the number of I/O operations by 2/3. In tables with very large numbers of columns and a large number of rows, storage efficiency is even greater.

A massively parallel processing (MPP) architecture parallelizes and distributes SQL operations to take advantage of all available resources. Machine learning delivers high throughput, irrespective of workloads or concurrent usage, using sophisticated algorithms to predict query run times and when queuing can begin.



## Data lake integration and extension

When data lakes first came on the scene, they were hailed for making it possible to store many different types of data for analytics. A modern data warehouse easily integrates with and complements data lakes, enabling the extension of queries to the lake without loading data, creating an insight bonanza with very little overhead. Put another way, the query processing layer links to files in a data lake that make the data in it immediately accessible—no data lift and shift, or copy required. Organizations can store highly structured, frequently accessed data on local data warehouse disks while keeping exabytes of semi-structured and unstructured data in the data lake. They can then query seamlessly across both to gain unique insights that would not be possible with independent datasets.



## Diverse types of data and broader set of analytics

Traditional, on-premises data warehousing uses ETL to extract data from a pool of data sources, holds it in a temporary staging database, and then transforms data into a form suitable for the target warehouse system. The structured data is then loaded into the warehouse, ready for analysis. This process locks data into proprietary formats, so they cannot be easily accessed other tools, without having to move the data. It also does not support open formats.

With a modern data warehouse, it is possible to access and query open file formats that organizations already use, such as Avro, CSV, Grok, JSON, ORC, Parquet, and more. Tools and interfaces enable SQL queries on data warehouse clusters, displaying the query results and query execution plan (for queries executed on compute nodes). As a result, data is easily accessible by a broader set of tools for analytics.



## Fully managed administration

A modern data warehouse automates some of the most time-consuming activities: backup and recovery, installing, configuring, patching, upgrading software, replication and more. This allows companies and organizations to focus on what really differentiates their organizations—such as analyzing petabytes of data, delivering video content, or building great mobile apps – and to leave the heavy lifting of the underlying technology infrastructure to AWS.

In addition, your data is automatically and continuously backed up to your data lake. The automation also includes asynchronously replicating your snapshots to a data lake in another region for disaster recovery. Your cluster is available as soon as the system metadata has been restored, and you can start running queries while user data is spooled down in the background.



## Security and data governance

Modern data warehouse security includes SOC1, SOC2, SOC3 and PCI DSS Level 1 eligible compliance and end-to-end data encryption that secures data in transit and data at rest. Plus, the modern data warehouse also offers protection against accidental or malicious data loss. If new data security threats emerge, the flexibility that is inherent in data warehouse modernization and the architecture of the warehouse enable quick design and implementation of new countermeasures.

For data governance, there is fine-grained, role-based access control for data and actions so that only people with the proper authorization can use and access data. Firewall rules can be configured to control network access to the data warehouse cluster or the warehouse can be isolated in an organization's own virtual network. The modern data warehouse also logs all SQL operations that can be accessed with SQL queries or downloaded to a secure location in the data lake, ensuring availability for audits and compliance.



## Pay as you go

With a modern data warehouse, you can pay as you go, starting small and scaling out to pricing by terabyte when you need it. As a result, modern data warehouses are less expensive than monolithic on-premises warehouses. You can analyze data faster, spinning up a cluster in just a few minutes, a major cost savings. Without incurring wait-time, resource, or downtime costs, you get higher performance and more scalability. That puts much less stress on procurement and planning.

# Why data warehouse modernization with AWS EPEXITY?



Amazon Redshift is a modern data warehouse solution from AWS that brings the above characteristics of the modern data warehouse together. With a large customer base, Redshift is the fastest growing cloud data warehouse, powering mission critical analytical workloads. Redshift extends data warehouse queries to your data lake, with no loading required. You can run analytic queries against petabytes of data stored locally in Redshift, and directly against exabytes of data stored in Amazon S3. It is simple to set up, automates most of your administrative tasks, and delivers fast performance at any scale.

EPEXITY has experience with multiple AWS services for implementing different workloads—from complex business intelligence workloads and ad-hoc and interactive queries to loading data to big data processing. When you combine our knowledge, offerings, and technical ability with Redshift, you are able to take advantage of the following benefits.



## Agility

Setting up an on-premises data warehouse can be a multi-million-dollar project that drains resources and time. By contrast, a modern data warehouse solution on AWS Redshift and implemented by APN Partners is easy to set up, deploy, and manage. In fact, you can deploy a new data warehouse in minutes. Or, for a modern data warehouse tailored specifically for your needs, APN Partners are available to architect, implement, and manage a fast and flexible platform.

When data warehouse automation and AWS Partners handle these time-consuming and labor-intensive tasks, you can focus more on your data and analytics. Plus, if you'd like help with the analytics, you have access to industry-leading tools and experts for loading, transforming and visualizing data, all available from AWS data integration partners.



## Breadth of functionality

Data warehouse modernization with Redshift and APN Partners offers a wide range of functionality. AWS delivers MPP, fault tolerance, compute, integration with third-party tools, limitless concurrency, audit and compliance, and network isolation. ETL is possible with Amazon Glue, or with APN Partner products. Other partner products can help you set up query and data modeling.

Because the open data format opens the door to a broad range of analytics capabilities, your organization can incorporate more sophisticated analytics. You are not limited by the types of analytics you use. Instead, you have the mechanisms for predictive, near-real-time, and prescriptive analytics and modeling along with business intelligence and other business analytics solutions and platforms.



## Global reach

With AWS availability zones all over the world, you gain resilience and uninterrupted performance, even during power outages, internet downtime, floods, and other natural disasters. As a result of this global reach, data warehouse modernization with AWS and our partners can offer the data sovereignty and local geographies for spinning up clusters that are inherent in AWS implementation. You can also place resources, such as instances, and data in multiple locations.



## Faster insights at a lower cost

Data warehouse modernization powered by AWS and APN Partners delivers insights at high speed. Local attached storage maximizes throughput between the CPUs and drives, and a high bandwidth mesh network maximizes throughput between nodes. Machine learning predictions of incoming query run times makes it possible to assign queries to the optimal queue for the fastest processing. Result caching delivers response times that are under a second for repeat queries, giving a significant boost to dashboard, visualization, and business intelligence tools like Quicksight.



## Data warehouse modernization your way

AWS delivers the platform you need to build a modern data warehouse. You have your choice of several solutions plus additional tools to set a firm foundation. And, when you work with APN Partners, you can rest assured that the partner has demonstrated success in helping customers like you who are either in your industry or have the same use cases. They understand your specific challenges and requirements and can evaluate and use the tools and best practices for collecting, storing, governing, and analyzing data—at any scale.

**EPLEXITY**

 [eplexity.com](https://www.eplexity.com)

 **888-501-5979**