



4 Imperatives for Your Big Data Analytics Journey



4 Imperatives for Your Big Data Analytics Journey

Harnessing the power of big data has never been more important for business success

In today's dynamic business and scientific environment, leading organizations rely on data analytics to differentiate their offerings, outperform competitors, and adapt quickly to meet their customers' needs. They see how the systematic use of data analytics and business intelligence reveals opportunities for better business outcomes and even new business models. Building a sustainable business for a long period of time isn't easy. It requires reinvention—likely multiple times over.

With any major organizational initiative, from creating new customer experiences to building new revenue streams, leaders must be able to quickly gather insights and get to the truth. They must be able to easily leverage data, no matter its source, size, or shape. Big data analytics along with artificial intelligence (AI) and machine learning (ML) helps businesses solve complex challenges and uncover new opportunities.

**“Only 50% of
businesses that
were on the
Fortune 500 list
in 2000 are still
there today”**

Reinventing your business starts with a strategy for centralized data access

Data, the lifeblood of all business, is pouring in from sensors, networks, applications, and an ever-expanding horde of connected devices. In 2020, every person generated 1.7 megabytes in just one second. But does your organization even have access to all its data or is it blind to the data that matters most? Are the teams that need that data waiting in queue or are they creating their own shadow copies and working from the best they can roughly assemble? Can they draw meaningful insights from quickly accessing the data at scale?

Centralized data access provides an easy way to define and manage security, governance, and auditing policies—all in one place. This enables your organization to provide fine grained access to data to the right user at the right time and effectively meet their regulatory governance and compliance requirements.

Data lives in silos within organizations, probably including yours. For example, marketing data may live apart from the product roadmap and strategic planning data. The key to a robust data and analytics strategy is having visibility to the data and connection and movement between the data sources, to get a 360-degree view of environments, markets, business and customers. With integrated access to data lakes, purpose-built data stores and databases through a common cloud architecture, you can relate PR and marketing data to sales, sales to customers and products and your customer responsiveness will improve. The cloud infrastructure for data management brings great agility, scalability, connection and cost efficiency.

When you centralize your data, you eliminate the silos and store it in a way that makes it easily accessible to your teams when they need it. Moving data is costly, time consuming and prone to errors. Data centralization is necessary to get a 360-degree view of your environment, market, business and customers. Without centralized access to of data, you can't relate PR and marketing data to sales, and sales to customers and products, and your customer responsiveness will suffer. Cloud is a requisite part of centralized data access due to its self-service model where you avoid infrastructure management. Other obvious benefits are connection, agility, scalability, elasticity and cost efficiency.

Centralized data access empowers your teams to make better decisions, improve customer experiences, and uncover new opportunities.

IT investment drivers

Head of IT

- 1 Data/business analytics
- 2 Security/risk management
- 3 Enterprise applications
- 4 Customer experience technologies
- 5 Machine learning/AI

Line of business

- 1 Security/risk management
- 2 Cloud migrations
- 3 Data/business analytics
- 4 Employee experience technologies
- 5 Data center/infrastructure

Tech Initiatives Driving 2021 IT Investments – 2021 State of the CIO Executive Summary



A cloud data warehouse is central to your big data analytics strategy

The data warehouse serves as a central foundation that enables activities like business intelligence, reporting, dashboards, and the other interfaces your team uses to best inform business decisions. It's similar but different from a data lake, in that data lakes allow you to store all of your data at any scale in open data formats without having to structure it. Whereas a cloud data warehouse enables you to quickly analyze small batches or very large chunks of data using standard SQL commands. A modern data warehouse is the fastest way for users to run analytic queries, power dashboards, and perform other mission critical SLA-bound operations. It can ingest data from different sources and formats and scale compute and storage independently to reduce overall costs. The right data warehouse should progress over time to deliver additional value to your organization and enable new use cases in a future-proof manner.

Cloud data warehouses offer the unique benefit of allowing organizations of all sizes to analyze large pools of variant data and extract significant value from it while keeping a historical record. They allow you to perform data analytics at any scale with cloud benefits such as agility, elasticity, operational savings and readily available applications your organization needs. Running a cloud data warehouse also connects you to a larger ecosystem of integrated, complementary cloud services, which engenders a more innovative environment for your organization. With a modern cloud data warehouse, you get enough flexibility to slice and dice data for closer examination whether at a high level or at a very fine, detailed level.

By 2022, 50% of new system deployments in the cloud will be based on a cloud data management ecosystem

Gartner

Solution Comparison for Cloud Data Warehouse, July 2021

Business imperatives for a modern data warehouse

When choosing a cloud data warehouse for your organization, consider four important attributes of data warehouses and how they will contribute to your business goals going forward.



Ability to analyze all of your data

Integrating a data lake, a data warehouse, operational databases and purpose-built stores gives you the power to analyze all of your data as a single source of truth while experiencing consistent performance. Leveraging AI/ML is necessary for deep business intelligence reporting, product innovation, revenue and profit prediction, and improving business operations.

The ideal approach is an integrated data platform with minimal data movement to quickly access, analyze it and act on it. Then you'll get the most out of the data while experiencing consistent performance. You also need the ability to setup up governance and compliance in a unified way to secure, monitor, and manage access to the data.



Easy for the entire organization to use

Whether the end-user is in the line of business or engineering, getting started, ingesting, accessing and analyzing data shouldn't be complicated. The user interfaces must be intuitive and highly functional to give you the right insights at the right time. Data sharing and collaboration with teams and partners while meeting your security and compliance requirements should be easy. Automation is essential to avoiding infrastructure management and realizing faster time to value. Software developers must have the ability to access data through an easy-to-use API, and to access semi-structured data like web data and IoT data without going through a laborious process.



Speed that keeps up with your business

Real-time streaming analytics enables you to analyze and process high volumes of fast streaming data from multiple sources simultaneously. You can discern relationships in information extracted from a number of input sources including devices, sensors, social media feeds, and applications, which can in turn be used to trigger actions for customer or organizational responses. A solution that delivers leading concurrency performance is a cornerstone of streaming analytics.

In order for your business to move quickly, data from all customer touch points, partnerships and business processes must automatically and reliably flow into a centralized location for deep, holistic analysis. This is where you derive value from it.



Integration with cloud services

Organizations like yours are looking to create the best environment to produce innovation. Part of that goal entails having a cloud data warehouse that integrates with other cloud services and partner solutions, resulting in a seamless user experience and better ROI. Incorporating machine learning can revolutionize your business by enabling a dramatically improved customer experience. An integrated service offering that comprises data lake and operational databases with centralized access gives you a single source of truth.

Organizations looking to create the best environment to engender innovation need a cloud data warehouse that integrates with other cloud services, like data lakes and AI/ML technologies, as well as and partner solutions that offer industry specialization or customization. Amazon Web Services (AWS) provides a natural environment for your business to grow quickly.

Amazon Redshift: A history of leadership in innovation

At AWS, we apply a ten-year history of innovation to deliver a modern data experience to our Amazon Redshift customers. We prioritize price performance leadership and enabling new use cases. Through global telemetry, we get an aggregate view of our customer workloads and use these insights to continuously improve performance. Focusing on the workloads that matter leads to incremental changes, which add up to substantial improvements over time.

The culture of innovation at AWS has led to a price performance flywheel. We believe consistency is a tenet of performance: You can scale your data volume on Amazon Redshift from 1 TB to over 1 PB with predictable cost and performance—and the all-important metric of price performance improves as your data scales.

Organizations of every size and industry use Amazon Redshift to process exabytes of data per day to power analytics workloads such as real-time business intelligence reporting, dashboarding applications, data discovery and scientific exploration, as well as streaming analytics. Amazon Redshift gives you an easy way to share data across internal or external accounts, while enabling secure and governed collaboration. Developers can easily build on top of it and access data in multiple formats. Even if you have no data warehouse experience, you'll find it easy to use, while experiencing the best price performance for all of your workloads without having to worry about infrastructure.

Amazon Redshift meets and exceeds the four key imperatives for a modern data warehouse. Read on to learn how it addresses each one.

**30% overall
analytics team
productivity
improvement,
408% 5-year ROI,
47% lower cost of
data management
platform**

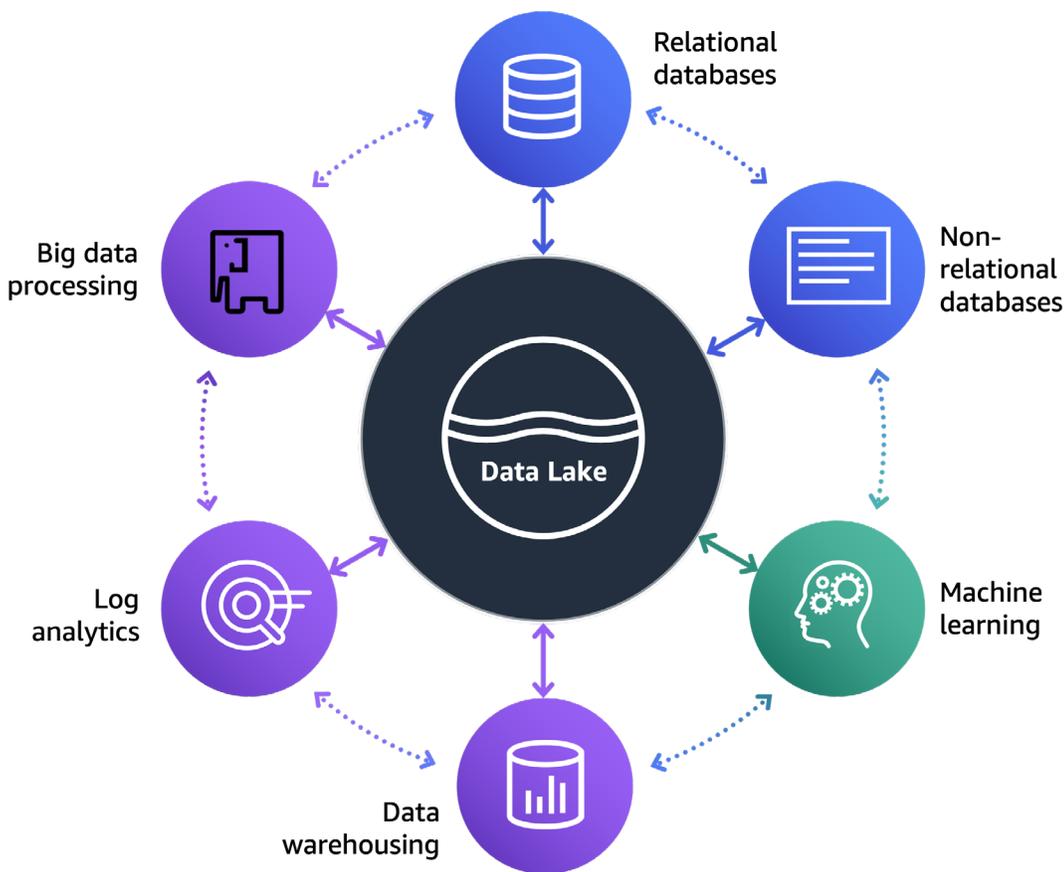
IDC: The Business Value of
Amazon Redshift Cloud Data
Warehouse - October 2021

Analyze all your data



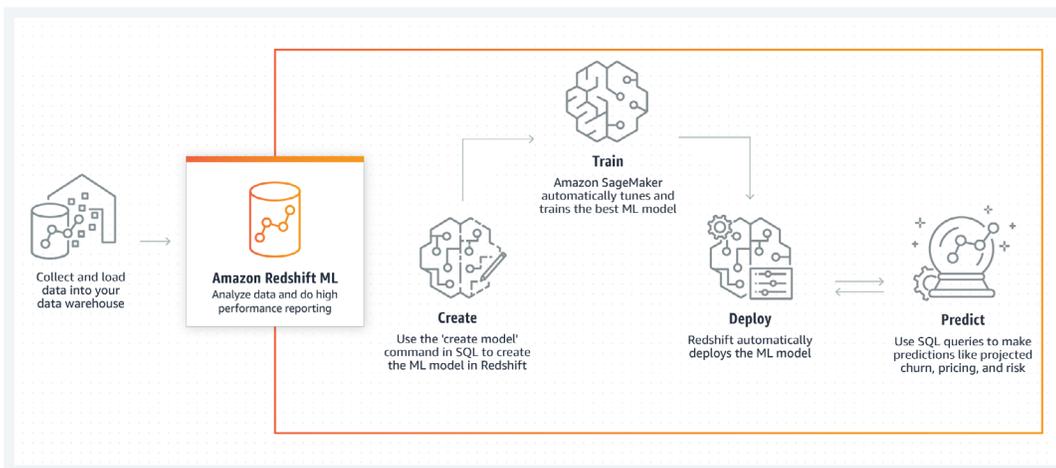
The amount of data generated by IoT, smart devices, cloud applications, and social is growing exponentially and requires a way to easily and cost-effectively analyze it with minimal time-to-insight, regardless of the format or where the data is stored.

Our unique approach enables you to connect any amount of data from various sources into an Amazon Simple Storage Service (Amazon S3) data lake. Unlike other cloud providers, organizations can store their data in S3 using standards-based open data formats to avoid being locked into any one proprietary data format or approach to analytics. Storing data in standards-based open formats makes it easy for any analytics or machine learning service to work on the data. It also eliminates the need to unnecessarily move, transform, or re-format the data in order to get value from it. Amazon Redshift efficiently extends your queries to your Amazon S3 data lake, making it easy to get value from it, and scales to petabytes and beyond.



Leveraging ML helps your entire organization imagine new products or services, transform customer experience, streamline your business operations, and improve decision making. Amazon Redshift is foundational to ML because it offers real-time streaming data for immediate insights, making it easy to extract value from all of your organization's data, no matter the scale. Amazon Redshift ML makes it easier for analysts and data scientists of all skill levels to use machine learning to transform the business.

Training ML models is time consuming and costly. We've made it easy with a simplified ML flow including integrated services, making it easy to create models for predictions. Redshift ML enables all kinds of new use cases, such as delivering personalized web experiences, forecasting demand, streamlining supply/demand decisions, or combining historical time series data with variables such as product features, pricing, and holidays. Other important uses include reduced customer churn, predictive maintenance, personalized recommendations, contact center modernization, improved safety and security, and increased customer engagement.



Analytics Teams

30% more productive on average

Productivity gains

Business analyst teams

34%

Data scientist teams

23%

Business intelligence teams

33%

Analytics engineer teams

29%

Analytics KPI benefits

71% more features added annually

62% higher query volume

27% faster to deliver reports to lines of business

IDC: The Business Value of Amazon Redshift Cloud Data Warehouse - October 2021



Zynga doubles ETL performance



Challenge:

Zynga develops some of the world's most popular social games, including Words with Friends, Zynga Poker, and Farmville, which are played by more than 70 million users every month. The company uses analytics to determine whether a game connects with its end users thereby supporting the company's mission. Zynga needed a partner to help it figure out how to meet the needs of its different games and scale for various stages of adoption.



Solution:

By migrating its data warehouse to Amazon Redshift, Zynga realized a dramatic performance improvement and was able to scale processing to terabytes per day while understanding the player experience and optimizing it.



Results:

- Improved gaming experiences, making games more social, interactive, and fun
- Consistently doubled extract, transform, and load (ETL) performance
- Easily scaled to process over 5.3 TB of game data generated each day



Jobcase recommends job search content at scale



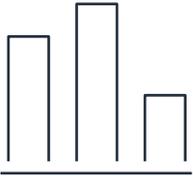
Challenge:

Jobcase connects millions of people to relevant job opportunities, companies, and other resources daily. The recommender system applies ML models to very big datasets, but the data and ML models aren't co-located on the same compute clusters, which requires the IT team to move large amounts of data across networks and build data pipelines. The data/model colocation issue creates a bottleneck for data scientists to perform quick experimentation and drive business value



Solution:

Using the in-database local inference capability provided through Amazon Redshift ML, Jobcase can perform model inference on billions of records in a matter of minutes, directly in their Amazon Redshift data warehouse. Amazon Redshift ML enables Jobcase to bring cutting-edge model classes with in-database local inference capabilities directly into Redshift and vastly increase the expressive power of the models.



Results:

- Effectively matched jobs to over 10 million active members on a daily basis
- Runs ML-based predictions at scale, performing billions of predictions in minutes
- Saves cost on external ML frameworks and compute with local in-database inference capability
- Improve member engagement 5-10%, resulting in increased revenue



Easy to run, share and collaborate



Being easy for the entire organization to use is a cornerstone of a modern data warehouse.

Productivity matters. We invested heavily in automation and features that work out-of-the-box to offload the routine busy work that holds back your organization.



Automation for ease of use

Automation makes data analytics easy. For example, Redshift's Auto workload management (WLM) feature with adaptive concurrency leverages ML to predict the resource utilization and runtime of each query. It works by dynamically predicting and allocating the amount of memory needed to run optimally with no investment and effort.

“By adopting Auto WLM, our Amazon Redshift cluster throughput increased by at least 15% on the same hardware footprint. Our average concurrency increased by 20%, allowing approximately 15,000 more queries per week now. All this with marginal impact to the rest of the query buckets or customers. Because Auto WLM removed hard walled resource partitions, we realized higher throughput during peak periods, delivering data sooner to our game studios.”

Alex Ignatius

Director of Analytics Engineering and Architecture for the EA Digital Platform

Electronic Arts, Inc., a global leader in digital interactive entertainment, realized immediate benefits from Amazon Redshift Auto WLM to gather player insights.

Another example is a self-performance-tuning capability called automatic table optimization (ATO) that helps you realize the best possible performance without manual effort. ATO uses ML to optimize performance for the workload, requiring no intervention.

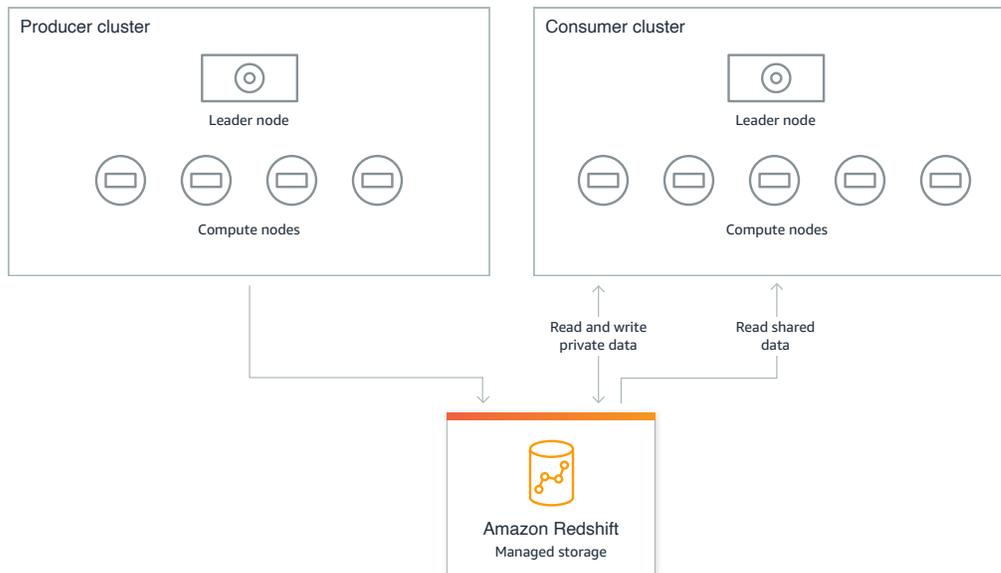
Amazon Redshift enables you to access and analyze data without worrying about tasks such as hardware provisioning, software patching, setup, configuration, or backups. It scales the underlying resources, enabling you to optimize your resource utilization and pay only for those resources.



Data sharing and collaboration

In a traditional data analytics model, your team follows a cumbersome process of manually unloading files from one system and copying them to another. This system fails to provide up-to-date views of the data because manual processes introduce delays and data inconsistencies.

Amazon Redshift data sharing provides instant, granular, and fast data access without copying data. You're able to query live data constantly, updating views across all organizations, customers, partners, and other third parties. Amazon Redshift shares data securely and enables governed collaboration with fine-grained access from databases, tables, views, and user-defined functions. Workloads run independently of one another, enabling your administrators to charge groups based on usage. This enables you to securely share live data with the same or different AWS accounts while tracking usage and retaining control of the data sets.



Faster software development

The goal of any software development team is faster iterations and time to market. In order to achieve that goal, your organization needs to efficiently process and query semi-structured or nested data such as web logs, data from industrial machines and equipment, sensors, genomics, and other sources.

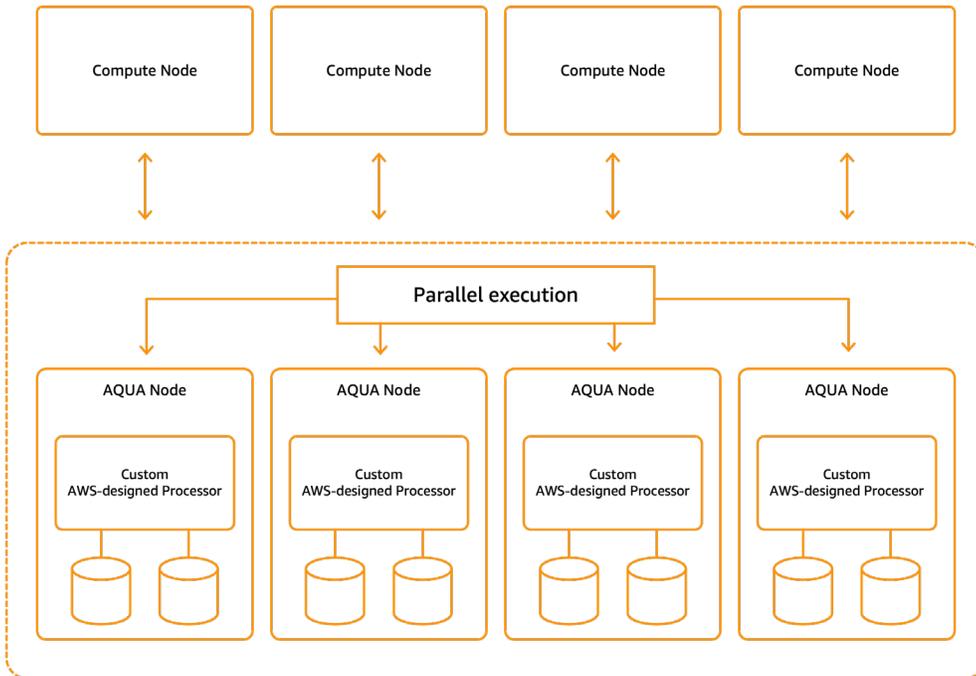
Amazon Redshift makes it possible to setup a DevTest environment quickly and easily without learning multiple complex functions or new third-party tools, because it allows your developers to access data in multiple formats natively. The easy-to-use Data API enables you to load data into Amazon Redshift and focus on building your applications without worrying about configuring drivers, database connections, or connection pools.

Faster time to value



Industry leading performance is required to stay ahead of the competition. Your organization needs fast access to data for decision making and business intelligence.

Amazon Redshift contains a number of innovations to deliver leading price performance. In fact, it's up to 3x better than other cloud data warehouses. We've improved query performance –up to 10x through features like advanced query accelerator (AQUA), which pushes the computation closer to the data. AQUA for Amazon Redshift works across large data sets and requires no code changes or extra cost.



AQUA Nodes with custom AWS-designed analytics processors make operations (compression, encryption, scans filtering, and aggregations) faster than with traditional CPUs. AQUA minimizes data movement over the network by pushing down operations to AQUA Nodes.

“FOX Corporation’s mission is to give millions of viewers the simple pleasure of being transported by a story on a screen. We have global audiences consuming premiere content across News, Sports, and Entertainment, and data is at the center of everything we do. Amazon Redshift empowers us to analyze petabytes of structured and semi-structured data across our data warehouse, operational database, and Amazon S3 data lake to discover, analyze, and activate data-driven decisions and powerful insights. As our petabyte-scale data continues to grow rapidly, we have been testing AQUA for Redshift to get better performance for our analytics queries while keeping our costs flat. We are seeing AQUA for Amazon Redshift improve the performance of some of our analytics queries by an order of magnitude and it is an example of how we are using latest technology to deliver a more personalized, curated, and timely experience to our viewers.”

Alex Tverdohleb

VP, Data Services - FOX Corporation

Amazon Redshift enables you to support virtually unlimited concurrent users and concurrent queries with consistently high performance. It allows you to perform real-time streaming analytics by processing high volumes of data from multiple sources simultaneously. To provide the best customer experience possible, we offer an hour of free concurrency scaling every day.

GE Renewable Energy increases wind energy production



Challenge:

Today's GE Renewable Energy wind turbines use sophisticated digital capabilities to collect data, run diagnostics, monitor production, and optimize the turbine as it operates. And GE needed a way to gather, monitor, analyze, and take action on all this turbine data—anywhere in the world.



Solution:

Using AWS services, GE has created a data lake where it collects and analyzes machine data captured at GE wind turbines around the world. GE relies on Amazon Simple Storage Service (Amazon S3) to store and protect its ever-expanding collection of wind turbine data and Amazon Redshift to gain new insights from the data it collects. These services also provide a foundation for building out AI/ML capabilities in the future.



Results:

- Increased energy production by as much as 20%
- Enables engineers to virtually monitor data at the farm or single turbine level
- Supports global access and coverage on with the world's most secure cloud

Weather reports, map directions, tweets with geographic positions, store locations, and airline routes rely on geometric (spatial) data to represent geographic features. Spatial data plays an important role in business analytics, reporting, and forecasting.

Amazon Redshift enables you to easily query spatial data, whether the data represents simple geometric objects such as points, lines, and polygons, or more complex structures like 3D objects, topological coverages, linear networks, and triangulated irregular networks.



Integration with cloud services



Amazon Redshift integrates with other AWS services to give you a seamless out-of-the-box experience that fosters innovation. Stream data into Amazon Managed Streaming for Apache Kafka (Amazon MSK), our fully managed service for Apache Kafka. Process the data through Amazon EMR, our big data platform for processing vast amounts of data using open source tools. Store and protect your data in Amazon S3, then load portions of it into Amazon Redshift, where users who need it can access it. From Amazon Redshift, you can then easily perform analytics at scale using fast queries to extract the most value from your data.

Use Redshift ML to generate predictions from a broad variety of data sources, and use Amazon SageMaker to create, automate, and manage end-to-end ML workflows at scale. SageMaker Pipelines helps you automate steps of the ML workflow, including data loading, data transformation, training, tuning, and deployment. Then, you can use Amazon QuickSight to easily create and publish interactive business intelligence dashboards that include ML-powered insights. Redshift also integrates with many partner solutions.

This is a small sample of the AWS service integrations that deliver industry leading price performance, scalability, and elasticity, while being easy to use. And best of all, everything works out of the box.



Conclusion

Today, modern cloud data warehouses are changing the way we leverage analytics to transform our businesses and environments. Common use cases include market analytics, AI/ML-based analytics, industry vertical analytics (including financial services, gaming, SaaS businesses and healthcare), as well as real-time analytics. As you embark on your big data analytics journey, look for a partner that offers industry-leading price performance, broad automation, a roadmap focused on continual innovation, integrations with complementary cloud services, and a robust ecosystem for innovation.

Learn more about Amazon Redshift ›

