



Whitepaper

# Redshift Deep Dive to Accredit a Well Architected Framework





## What is AWS Well-Architected Framework?

AWS well-architected framework provides a consistent approach to architectures through number of questions based on six pillars — operational excellence, reliability, sustainability, cost optimization, security, and performance efficiency to check if architecture is affiliated to AWS best practices

## What is Redshift?

Amazon Redshift is a fully managed service, petabyte-scale data warehouse provided by AWS. Redshift data warehouses provide massively parallel, columnar, and ANSI SQL data warehouse that is easy to use. Redshift shines in heavy workloads and with strong integration with AWS services. It requires 5 min of tech team's time to create a cluster, after which intelligent machine learning based cluster tuning and intelligent automated scaling enables less

hands-on administrative effort. Business users can start using Redshift data warehouse within minutes of creation through user friendly and collaborative web-based editor v2. Redshift provides Redshift ML for enabling citizen data scientists to quickly deliver machine learning models using AutoML.

In this article, we will see how Redshift features are aligned with well architected framework and other important criteria.



Figure 1: Well architected framework pillars

## Security

Keeping AWS Redshift data inside customer VPC (private subnet) makes sure customer data is fully secured. Data encryption is achieved using AWS key management system. Access management is done by identity access management (IAM) service. Redshift provides features like cluster encryption, row level access control, column level access control. For example, single sign on for the customer email can be turned on (supporting their workflow for request, review and approval) and following this IAM roles can be used to manage roles such as cloud engineer, data analyst, and data scientist. Amazon Redshift uses hardware accelerated SSL to communicate with S3 or DynamoDB for UNLOAD, backup, and restore operations to protect your data in transit within the AWS cloud

## Performance Efficiency

Redshift provides customers the flexibility to fine tune their clusters to meet their workload demands, to get to a better performance optimizing the same compute resources, without incurring any additional cost. When benchmarked against top competitors in market, Redshift provides up to 3x price performance. It uses hardware accelerated engine called AQUA, and serverless fleet of spectrum nodes to performantly query data in Redshift and data lake. Redshift attains tremendously fast query execution by engaging these performance features.

- MPP(Massively parallel processing)
- Columnar data storage
- Data compression
- Query optimizer
- Result caching



## Operational Excellence

Amazon Redshift has latest UI for Amazon Redshift management and intelligent-automated everyday tasks for near zero administration. Powerful easy to use web query tool on AWS console only for all users. Redshift has super data type for native semi-structured, nested JSON data format, native spatial support, and postgres geo spatial support through federation. It also has NoSQL, Python, PL/SQL support. Amazon Redshift offers a web based analytics workbench - query editor v2.0 where users can run SQL queries directly against a cluster, save and export results for quick visualization and collaborate by saving queries organized in folders for easy access. This feature provides a one stop for easy access, quick visualization and collaboration across teams. This lets the teams focus on performing analysis to drive innovation without having to worry about administration of additional tools for access and collaboration. AWS Lake Formation service can be leveraged for governance activities and Amazon Glue can be used for data governance with basic features. Redshift works seamless with S3, Glue, Sagemaker Canvas, Quicksight which makes overall analytics journey easier for customer. It also seamlessly integrates with popular external business intelligence and ETL tools over ODBC/JDBC

## Sustainability

By using Amazon Redshift data sharing feature, many departments of organization can optimize their implementation for sustainability, avoiding superfluous storage and reducing data transfer between AWS regions. Data sharing enables instantaneous, granular, and fast data access across Amazon Redshift clusters without the need to copy or move it.

## Reliability

AWS provides Out of the box multi-AZ with zero RPO availability. Redshift continuously monitors the health of the cluster and repeatedly re-replicates data from unsuccessful drives and replaces nodes as necessary for fault tolerance. Clusters can also be displaced to another availability zones (AZs) without any data loss or application changes. Amazon cloudwatch, amazon cloudtrail services get easily integrated with redshift to monitor redshift resources. Monitoring is an important part of maintaining reliability of overall solution. Automatic snapshots with cross-region replication and restore points helps in data recovery. Amazon Redshift automatically takes incremental snapshots that track changes to the . Automated snapshots retain all of the data required to reinstate a cluster from a snapshot. You can create a snapshot schedule to control when automated snapshots are taken, or you can take a manual snapshot any time.

## Cost Optimization

Unlike alternatives where advanced security features require higher level of subscription, Redshift offers comprehensive security features in same level. With Redshift (scheduled or manual) pause and resume, and provisioned reserved instance clusters, you get great cost savings for both occasional use cases and regular use data warehouses. Amazon Redshift comes with Redshift Advisor which provides recommendation for cost optimization and performance tuning of the cluster. The advisor analyzes your cluster's workload to find the most appropriate distribution key and sort key based on the query patterns of your cluster.



## Other Features Redshift Provides

- **Modern Analytics Platform** - It is truly modern, easy to use and scalable analytics platform:
- **Automatic Scaling** - Handle virtually unlimited users with same query performance using concurrency scaling. Automatically scales up and down based on query volume. Allows fine grained cost control to avoid bill shock (unlike alternatives).
- **Redshift Serverless** - Redshift brings the flexibility of skipping managing data warehouse infrastructure to run and scale analytics. As per need, customer can switch between serverless and provisioned any time by backup and restore features.
- **Intelligent Machine Learning Auto Admin** - It automatically optimizes data warehouse without needing administrators to manually tune it, using machine learning models (auto table optimizations, auto vacuum, auto sorting, and auto materialized views).
- **Redshift ML** - Strong machine learning capabilities and integration with SageMaker - supports regression, classification and clustering unlike alternative which only supports regression and classification.
- **Great Collaboration Tools** - Web based editor (v2) allows analysts to create SQL notebooks, scripts, upload csv/json files, create chart visualizations, share with team members etc. to provide highly collaborative and productive modern analytics platform.
- **Strong Orchestration Options for ETL** - Step functions and managed Apache airflow.
- Change data capture from databases using no code DMS service, with great built-in reporting for CDC and copying of data.
- **Redshift Data Sharing** - Easy and secure way to share data across multiple Redshift clusters with high performance
- Stored procedure is in SQL instead of JavaScript, keeping analytics logic in same language.



## New Announcements in AWS RE:INVENT 2022

### **Amazon Redshift Integration for Apache Spark**

With this new release, it will be easy to build and run Spark applications on Amazon Redshift and Redshift Serverless, empowering clients to explore data warehouse for a broader solutioning in Machine learning and AWS analytics.

### **Amazon Redshift Streaming Ingestion**

With this new release, Amazon Redshift can natively ingest huge data from Amazon Kinesis Data Streams and Amazon MSK into an Amazon Redshift materialized view and query it quickly.

### **AWS Backup Support**

AWS backup service can support redshift now with this release. We can now protect redshift clusters using AWS backup

### **Informatica Data Loader for Amazon Redshift**

A new feature that allows customers to run high-speed and high-volume data loading to Amazon Redshift for free.

### **Auto Copy Support from S3**

Without using any new tool or custom solution, now we can setup continuous file ingestion rules to track S3 paths and automatically load new files.

## The Right Choice of Delivery Partner to Implement Redshift

With new features like Redshift ML, Redshift data sharing and data exchange, Redshift best fits in modern cloud stacks for the customers who want to scale their analytics platform for better performance. Recently Tech Mahindra is recognized as Amazon Redshift service delivery partner

For any Redshift related implementation, please contact us through:

<https://www.techmahindra.com/en-in/data-platform-modernization-with-aws-redshift>



## Authors



### Satyawan Balu Kadalag

*Principal Solution Architect,  
Data on cloud competency, Tech Mahindra*

Satyawan Kadalag is Principal solution Architect at Tech Mahindra having total IT experience of 17 years. He's responsible for the consultation and design of customers' cloud solution architectures for customers across globe. He has extensive experience in legacy and cloud databases.



### Sathish Arumugam

*Sr. Partner Solution Architect,  
Amazon Web Services*

Sathish Arumugam is a Sr. Partner Solution Architect at Amazon Web Services. Sathish is a Containers TFC AoD and AWS Data Analytics Specialty certified Solutions Architect. He helps partners and customers with the AWS Well Architected best practices in their cloud transformation journey and the business critical workloads hosted on the AWS cloud. In his spare time, he loves to spend time with his family and pursue his passion for Cricket



### Piyush Patra

*Partner Solution Architect, Amazon  
Web Services*

Piyush Patra is a Partner Solutions Architect at Amazon Web Services. He helps partners with their Analytics journeys supporting them for achieving differentiation, enabling their technical and sales teams on AWS native analytics services, setting up partners for success by helping design solutions adhering to recommended best practices. In his spare time, he loves to cook for friends and family and explore different cuisines



### Nikhil Sontakke

*Solution Architect, Data on Cloud  
Competency, Tech Mahindra*

Nikhil Sontakke is Solution Architect at Tech Mahindra having total IT experience of 11 years. He help customers across the globe for their cloud modernization journey on Hyperscalers, helps in understanding their business pain points and consulting/designing cloud based solution architectures. He has extensive experience in Data analytics and Data engineering.