

Implementing A Data Loud Utility (DLU) Pipeline with Amazon Redshift



MEET THE CLIENT

AARP is a nonprofit, nonpartisan social welfare organization with a membership of nearly 38 million that helps people turn their goals and dreams into real possibilities. It does so by building communities and advocating for the most critical problems for families, including healthcare, education, income stability, and financial abuse prevention. AARP provides discounts and rewards to its subscription members in different healthcare and other lifestyle categories.

AARP Services Inc. (ASI) handles the partnerships with organizations offering benefits to AARP members. The firm studies trends in that segment, including attitudinal patterns, behavioral tendencies, purchasing trends and more. It has strategic partnerships with other businesses, offering valuable and relevant insights that include research on the 50+ age segment with positive communication channels for a fee.

THE CHALLENGE

Every day, businesses make important business decisions to better understand and improve operational performance. The process of exploring that data and analyzing reports on-demand to extract meaningful insights has become the biggest challenge for AARP using conventional OLTP (Online Transaction Processing) systems. Due to this, the company wants to load the data on-demand/daily into managed DWH's (Data Warehouses) with automation in place and warning mechanism in case the given threshold exceeds it.



Find out how Iexcel solutions can help your business. Contact us today!

inquiry@idexcel.com | www.idexcel.com

Management & Maintenance Challenges

- **Less Flexibility:** With conventional OLTP, systems cannot be scaled up on demand as per the data load.
- **Lack of Resources:** Management and maintenance require recruiting the right people who have experience with the right tools.
- **Project Timeline Concerns:** Complexity can create challenges leading to lengthy and difficult deployments.

Technical Challenges

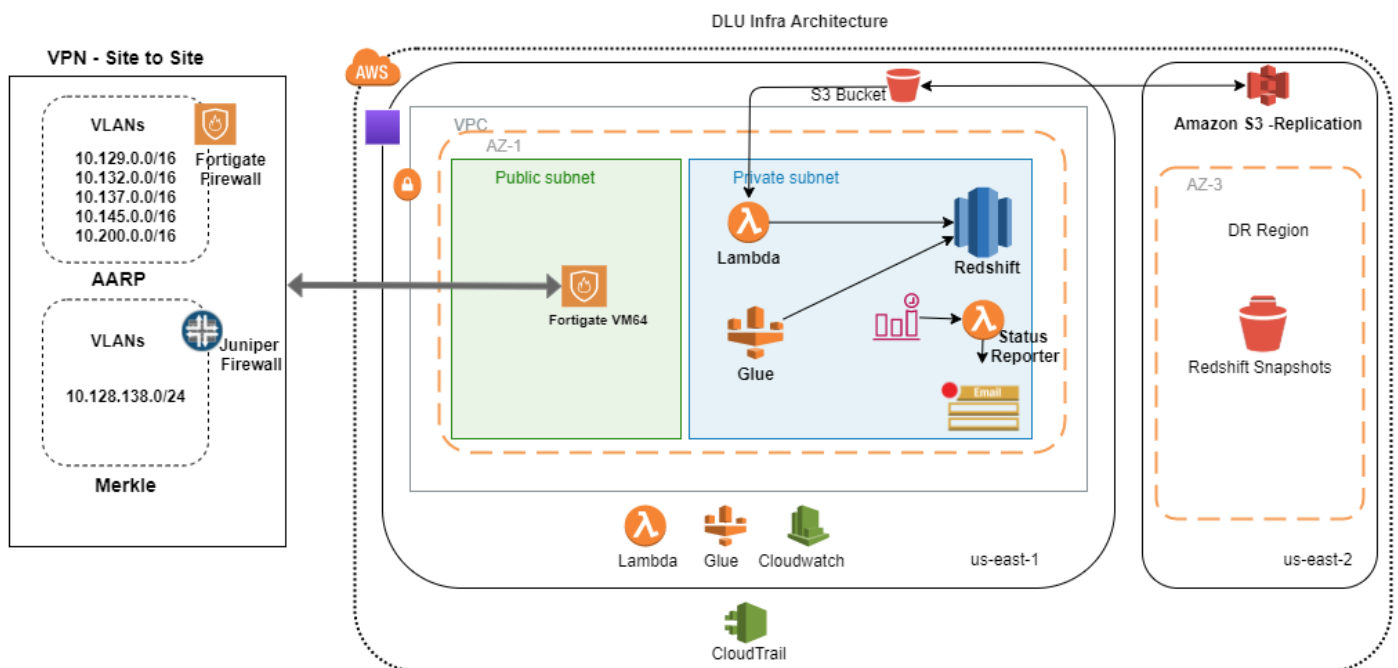
- **Operational:** The current solution lacks strong data processing and query performance and impacts the leadership team's ability to make decisions at the right time.
- **Performance Issues:** Maintaining optimal performance while running complex SQL queries.
- **Efficiency:** Disk I/O is slow.
- **Flexibility:** Building high-level architecture best practices that are secure, scalable, and resilient.

THE IDEXCEL SOLUTION:

To solve the above challenges, Idexcel had designed a cloud-centric data load utility pipeline. In our use case, the target database is Redshift Data Warehouse where the data load is triggered through events using many AWS managed services. It operates in two modes for Redshift DB:

1. **Append:** In this mode, data will be appended to the existing table in Redshift Data Warehouse. This mode is typically useful for datasets that are not often updated.
2. **Overwrite:** In this mode, the existing Redshift table is truncated, and the newly processed data is loaded into the table.

The entire pipeline is fully automated and audited in each and every step for transparency in the process and easy debugging. Also, a summary report will be generated and sent to clients that summarizes all the data load jobs in the agreed format.



Find out how Idexcel solutions can help your business. Contact us today!

inquiry@idexcel.com | www.idexcel.com

IDEXCEL'S SOLUTION INCLUDED:

1. AWS Glue service for ETL process i.e., loading data from S3 to Redshift using lambda to initiate the ETL processes.
2. Infrastructural health monitoring with efficient notification system using CloudWatch, SNS solution.
3. Amazon Virtual Private Cloud (VPC) and its associated services for securely isolating cloud resources and S3 data is secured using AWS default encryption (AES-256).
4. AWS CloudFormation to deploy infrastructure as code.

IDEXCEL'S SOLUTION USES THE FOLLOWING AWS SERVICES:

AWS Service	Purpose
S3	To store and retrieve any amount of data, at any time, from anywhere on the web.
Lambda	Serverless compute service that runs your code in response to events.
Glue	Managed ETL service that automatically scales underlying resources and retries jobs if they fail.
Redshift	Fully managed petabyte-scale cloud-based data warehouse product designed for large scale data set storage and analysis.
VPC	To build a virtual network in the AWS cloud.
Subnets	The subnets have been designed using CIDR blocks and segregated according to the purpose, e.g., App & DB and utility.
VPN	The VPN connects the AWS VPC to the customer network.
CloudTrail	To allow complete insight of the cloud environment.
Secret Manager	To protect DB credentials.
Trusted Advisor	To monitor the environment to validate the AWS Well Architected Framework Compliance.
CloudWatch	To monitor and alert any changes to some of the primary services of the entire infrastructure.
IAM Roles & Policies	To provide necessary permissions to Lambda code and accessibility of the S3 buckets.
Systems Manager	To execute various ad-hoc necessary PowerShell commands without logging to the instances.
Certificate Manager	Certificate obtained from 3rd party authority was imported into ACM.



Find out how Idexcel solutions can help your business. Contact us today!

inquiry@idexcel.com | www.idexcel.com

WHY DO WE USE REDSHIFT DATABASE SERVICE TO STORE MIGRATED DATA LOAD FROM ANY CONVENTIONAL DATABASES?

Amazon Redshift enables efficient storage and optimum query performance through a combination of massively parallel processing, columnar data storage, and targeted data compression encoding schemes. The data in the Redshift data warehouse is updated in real-time or at a frequency of the customer's choice. Amazon Redshift is a core solution for our customer challenges of optimizing disk space and increasing performance efficacy. With columnar data storage, massively parallel processing (MPP), and efficient integration with other AWS services, Amazon Redshift provides a very reliable and secure solution to business intelligence and analytics platform needs. This robust solution addresses data-driven enterprise goals to help them in their Digital Transformation journey.

KEY ADVANTAGES OF THIS SOLUTION

With proper assessment, planning and design, the Idexcel team was able to complete the migration process within the original SLA's/timeframe.



- **Cost Optimization:** Overall cost was reduced by using more AWS managed services and the infrastructure and maintenance cost was reduced using serverless services.



- **Scalable:** Supporting new workloads and increasing data makes this solution highly available, resilient, and flexible to meet business needs.



- **Easy Configuration & Management:** One-time setup and automated glue job alerts mechanism in case of job failures.



- **Automated Maintenance:** Redshift automates infrastructure provisioning and administrative tasks such as backups, vacuum, replication, and patching. Also, there are automated alerts from Redshift cluster if a defined threshold is crossed.

Interested in learning more about loading bulk amount of data to better manage your IT ecosystem?

[Contact Idexcel](#) to schedule a workshop, request a demo, or to speak with someone from our team about how we can help implement this solution.

OUR AWS COMPETENCIES



- Public Sector
- Solution Provider
- DevOps Services Competency

- Financial Services Competency
- Migration Services Competency

Contact us Idexcel, Inc.

459 Herndon Parkway Suite 10, Herndon, VA 20170
Tel: 703-230-2600
Email: inquiry@idexcel.com



Find out how Idexcel solutions can help your business. Contact us today!

inquiry@idexcel.com | www.idexcel.com