nClouds | AWS Case Studies
CyberCube

How nClouds helped CyberCube accelerate data analytics, reduce data prep time, and reduce costs.

CyberCube

Industry
InsurTech, Software

Location
San Francisco, CA

Challenge
Accelerate data analytics, reduce data prep time, and reduce costs.

Featured Services
AWS Glue, Amazon Athena

About CyberCube

CyberCube is focused on solving the most difficult and important cyber risk challenges in insurance with world-class analytics.

Their team is comprised of multi-disciplinary experts across data science, cybersecurity, software engineering, actuarial modeling, and commercial insurance. CyberCube offers a software-as-a-service (SaaS) platform for cyber risk aggregation modeling and insurance underwriting.

The platform was established in 2015 by Symantec to apply the cybersecurity company's unique sources of data, intelligence and expertise to cyber insurance analytics, and now operates as a standalone company. To learn more, go to https://www.cybcube.com/

Benefits Summary

Accelerated data analytics	Reduced data prep time	Cost savings

CHALLENGE

Challenge: Accelerate data analytics, reduce data prep time, and reduce costs.

CyberCube wanted to support its strategic growth by enabling their analysts to build insights that deliver additional value to their customers via new services on their cyber risk analytics platform.

They needed a solution that could easily scale to their size requirements in a transient model, handle data of virtually any type, store massive amounts of data at a low cost, and extract basic models from the data itself to create the ability to query without a lot of transformation.

CyberCube's existing data processing routine inspected data sources and hard-coded data around them. nClouds recommended a solution using AWS Glue and Amazon Athena to more easily structure and load the data for analytics, saving data preparation time.

> " As a result of our collaboration with nClouds, we can offer our customers new insights and services. And, we can deliver analytics results faster and at a lower cost. All of this is important to CyberCube's continued growth."
>
> **— Ajay Garg**
> **Head of Engineering, CyberCube**

"Similar to the needs of many data-intensive companies, CyberCube asked nClouds to help them reduce the amount of time spent by their data scientists searching for and preparing data," said John Jones, CTO, nClouds. "A 2019 survey by IDC found that 48% of data workers' time is spent on searching for and preparing data before any analysis or machine learning."

Why AWS and nClouds

CyberCube was an existing customer of nClouds, a Premier Consulting Partner in the Amazon Web Services Partner Network (APN). nClouds had helped CyberCube rebuild workloads with infrastructure-as-code, build a fully automated CI/CD pipeline, and follow best practices for security, cost, performance, reliability, and operational excellence.

Then, based on their excellent experience partnering with nClouds, CyberCube extended nClouds' work with them to include a project aimed at gaining expanded use of their data to support innovation and provide additional value to their customers.

nClouds applied their AWS technical expertise in data and analytics to help CyberCube more easily ingest large amounts of external data, push it into relational structures, and make immediate use of it without requiring a lot of administration bandwidth to create new Amazon EMR clusters each time processing is needed.

aws partner network

Premier
Consulting
Partner

DevOps Competency
Marketplace Seller
Migration Competency
MSP Partner
Well Architected

CyberCube leveraged several Amazon Web Services:

- **Amazon Athena** - An interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL. Athena is serverless, so there is no infrastructure to manage, and CyberCube pays only for the queries that are run.
- **Amazon EMR** - A cloud-native big data platform that enables CyberCube to process vast amounts of data quickly and cost-effectively at scale.
- **Amazon RDS** - Enables CyberCube to easily set up, operate, and scale a relational database in the cloud.
- **Amazon S3** - A flexible way to store and retrieve data, providing CyberCube with cost optimization, access control, and compliance.
- **AWS Glue** - A fully managed extract, transform, and load (ETL) service that makes it easy for CyberCube to prepare and load their data for analytics.

CyberCube's solution stack also included an additional, essential third-party tool:

- Apache Spark, a unified analytics engine for large-scale data processing.

nClouds' Solution Architecture for CyberCube

CyberCube engaged with nClouds to create a data processing pipeline that enabled easier access to their data to build insights that deliver new value to their customers.

nClouds implemented AWS Glue ETL service to prepare and load CyberCube's data for analytics. AWS Glue runs ETL jobs in an Apache Spark serverless environment. This service structures different formats of datasets in a common way and puts a metadata layer on top of it (the AWS Glue Data Catalog) so that all data can be accessed in the same way.

The AWS Glue Data Catalog provides the capability to crawl and create schemas for the complex data sources allowing them both to be easily ingestible for ETL and accessed using SQL via Athena. AWS Glue also provides the benefits of Amazon EMR compute power and scalability without the hassles of maintaining templates to create, start, stop, and run scripts in the Amazon EMR cluster.

nClouds used a crawler to populate the AWS Glue Data Catalog with tables (metadata definitions that represent the data in a data store) for use in ETL routines. The tables sit on top of Amazon RDS instances with Java Database Connectivity (JDBC) connections to the Apache Spark environment. JDBC connectivity supports most of the visualization solutions on the market.

Once the data is in the AWS Glue Data Catalog, it is queryable via Amazon Athena (which is out-of-the-box integrated with AWS Glue Data Catalog). Amazon Athena enables CyberCube to create a unified metadata repository across various services, crawl data sources to discover schemas, populate the AWS Glue Data Catalog with new and modified table and partition definitions, and maintain schema versioning. SQL can query that data as if it's a table in a database regardless of the underlying data structure, in preparation for analysis by a business intelligence tool.

High-level architecture diagram:
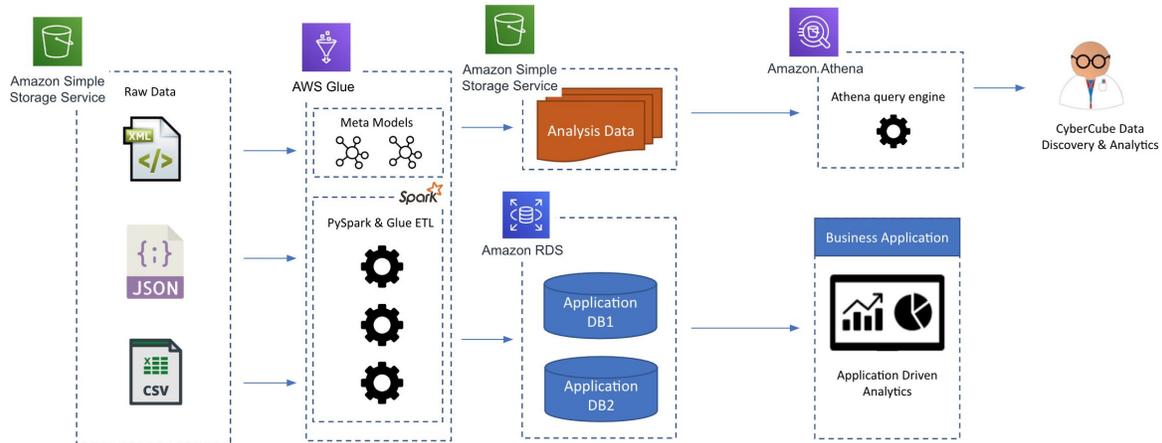
The Benefits

Teaming with nClouds, CyberCube now has a modernized, efficient, cost-effective data analytics system. This phase of the project has yielded numerous benefits:

Accelerated data analytics

Using AWS Glue, CyberCube's data is immediately searchable, queryable, and available for ETL. It automates much of the effort in building, maintaining, and running ETL jobs. AWS Glue is serverless, so there is no infrastructure to provision or manage. And, Amazon Athena scales automatically—executing queries in parallel—so results are fast, even with large datasets and complex queries.

Reduced data prep time

AWS Glue automates the time-consuming data preparation process. It takes datasets in different formats and converts them into a single, query-optimized format that can be consumed quickly by various analytical tools.

Cost savings

With AWS Glue, CyberCube will pay only for the time their ETL job takes to run – an hourly rate based on the number of data processing units (DPUs) used. There are no resources to manage, no upfront costs, no charges for startup or shutdown time. Likewise, with Amazon Athena, CyberCube pays only for the queries they run.

About nClouds

nClouds is a certified, award-winning provider of AWS and DevOps consulting and implementation services. We partner with our customers, as extensions of their teams, to build and manage modern infrastructure solutions that deliver innovation faster. We leap beyond the status quo.