Novatiq Upgrades Data Infrastructure, Scalability with Amazon Neptune Graph Database

Novatiq is a privacy-first technology platform that allows businesses such as telcos and publishers to drive more value from their data whilst making it safe to use across the advertising ecosystem. Built on proprietary patented technology, Novatiq's in-network telco solution is designed to enable publishers to verify users and help advertisers and brands to activate first-party audiences.

Novatiq has grown quickly since its founding and expects demand to increase significantly in the coming years. However, the company wasn't sure its Redis-based data infrastructure was the right approach to manage complex, multi-dimensional relationships while still meeting performance standards. Novatiq decided to bring in ClearScale, an Amazon Web Services (AWS) expert with extensive big data experience, to come up with a better solution.

"We were introduced to ClearScale via our ongoing relationship with Amazon's AWS Startups division. From the outset, ClearScale demonstrated significant experience in the data ecosystem and rapidly circled around our requirement to understand our objective. Their focus on enterprise grade delivery excellence was evident immediately, from commercial to project management, architecture, engineering, validation and handover, the engagement was flawless and delivered with impressive accuracy and productive communication at all times. It goes without saying that they delivered everything we asked for, so we will gladly work with ClearScale again."

Mark Stewart, Chief Technology Officer, Novatiq

The Challenge

Novatiq previously managed relationships between buy side of the advertising ecosystem, publishers and telcos through a series of Redis tables. While the approach worked in the past, Novatiq questioned whether its Redis-based solution would suffice going forward. The data structure store would have to scale tremendously over the next three years to meet estimated platform demands, which it wasn't capable of doing effectively.

Novatiq began exploring alternatives and determined the best solution would involve implementing a graph database to manage a vast number of data relationships. Furthermore, the Novatiq team wanted to take advantage of decoupled serverless architecture to process both on- and off-network requests within the core application. Novatiq also envisioned its new solution supporting analytics and consent requests.

Novatiq leaders decided to reach out to ClearScale for guidance. ClearScale has the AWS Data and Analytics Competency, which means the consultancy has a track record of executing big data projects successfully on the cloud. Given Novatiq's ambitious growth targets and performance requirements, ClearScale was the perfect candidate to assist.

The ClearScale Solution

Novatiq had three deliverables for ClearScale:
- Configure a data processing pipeline configuration
- Configure a reporting pipeline
- Optimize and test the platform for performance

With these goals in mind, ClearScale got to work.

Configuring the Data Processing Pipeline

For the data processing pipeline, ClearScale used an Application Load Balancer as an entry point for requests and implemented Lambda functions containing application logic. The team also relied on Amazon SQS queues to enable integration between the pipeline, the future graph database, and other systems.

ClearScale used Amazon Neptune, a fully managed graph database engine built specifically for the cloud, in place of the legacy Redis-based solution. Amazon Neptune can handle billions of relationships between data points and conduct queries with milliseconds latency, which is exactly what Novatiq needed. As a managed product, Amazon Neptune also takes on typical database management tasks, like hardware provisioning and software patching, so that Novatiq's engineers can focus on higher-value activities as the company grows.

Configuring the Reporting Pipeline

On the reporting pipeline side, ClearScale landed on the combination of AWS Glue Data Catalog and Amazon Athena. AWS Glue is a serverless data integration service designed to help engineers prepare data for advanced analyses and app development. Amazon Athena is an interactive querying service that allows data scientists to run analyses using standard SQL. In addition, Amazon Athena integrates with AWS Glue Data Catalog out of the box, making it easy for users, like Novatiq, to maintain a unified repository of metadata across different services.

Optimization & Performance Testing

After setting up the pipelines, ClearScale moved forward with performance testing and optimization. ClearScale's engineers first generated a dataset the size of Novatiq's one-year utilization expectation. Then, ClearScale re-wrote and optimized every query with Gremlin before using a Jupyter Notebook to evaluate performance.

Next, ClearScale tested the scalability of Novatiq's new pipelines and database. The team took a target requests per second (RPS) value from NFRs and built out a set of scenarios for load testing with concurrent execution. On top of that, ClearScale changed the structure of the dataset, along with Neptune cluster reader and writer sizes. The team ran a load test using a solution built on JMeter that produced the following results:

Test Dataset	RPS	Writer	Reader	Error Rate
10% updates	3000	db.r5.4xlarge	db.r5.4xlarge	0.15%
10% updates	6000	db.r5.4xlarge	db.r5.4xlarge	23.34%
10% updates	3000	db.r5.4xlarge	2 x db.r5.4xlarge	0.00%
10% updates	5000	db.r5.4xlarge	2 x db.r5.4xlarge	0.00%
10% updates	5000	db.r5.4xlarge	db.r5.4xlarge	1.58%
10% updates	6000	db.r5.4xlarge	2 x db.r5.4xlarge	0.00%
20% updates	3000	db.r5.4xlarge	2 x db.r5.4xlarge	0.00%
20% updates	3000	db.r5.4xlarge	2 x db.r5.4xlarge	0.00%
10% updates	3000	db.r5.2xlarge	2 x db.r5.2xlarge	0.00%
10% updates	5000	db.r5.2xlarge	2 x db.r5.2xlarge	0.00%

Using these results, ClearScale was able to define an optimal configuration based on anticipated workloads, create a scalability roadmap, and optimize code to reduce the overall error rate.

Architecture Diagram

Novatiq Off Network Consent Managment
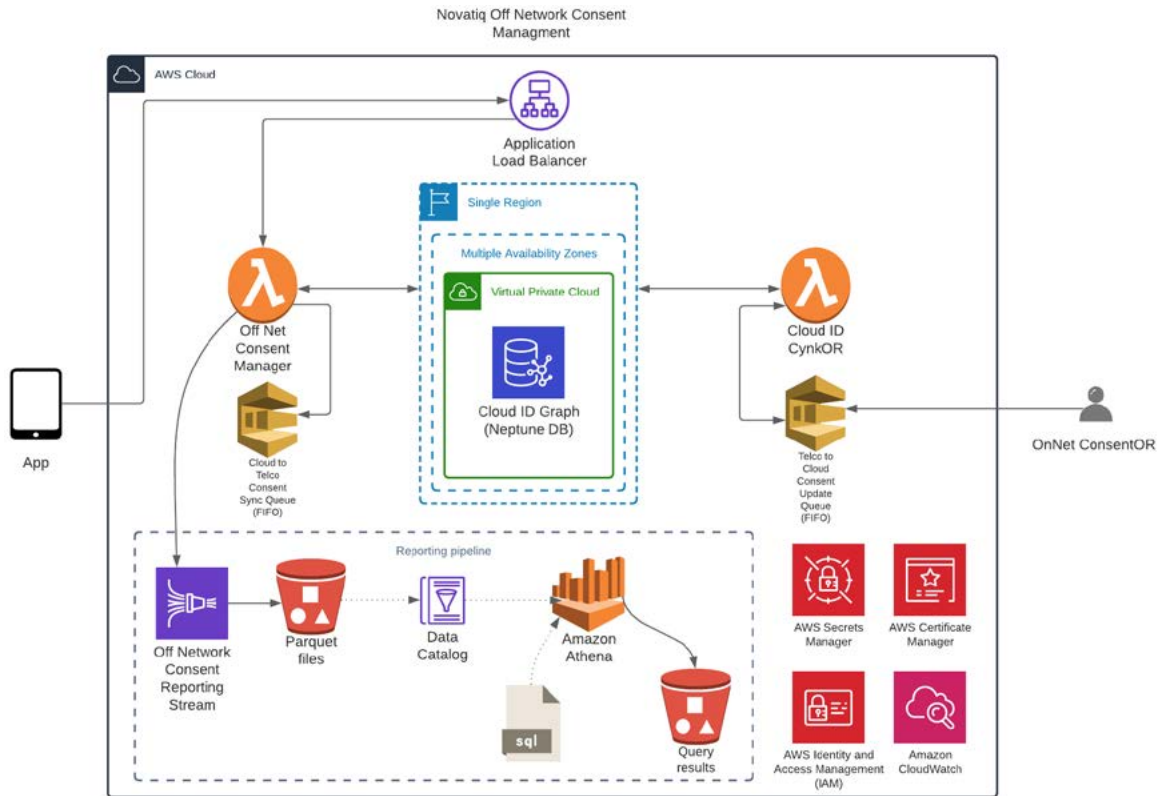
The Benefits

Thanks to ClearScale's support, Novatiq gained a secure and reliable graph database solution that can scale with the company's workloads going forward. The new processing and reporting pipelines can handle massive data volumes without hindering performance for Novatiq's end users. The pipelines are also built using fully managed AWS solutions, saving Novatiq engineers from having to spend too much time maintaining the company's IT infrastructure.

In addition, the scalability roadmap that ClearScale created gave Novatiq's leaders a clear understanding of what configuration adjustments will need to be made down the line to continue meeting performance targets. ClearScale also set Novatiq up with a suite of cost-effective solutions so that the business doesn't spend more than it needs to on its data platform.

At a time when big data capabilities and regulatory compliance are paramount, Novatiq's privacy-first platform does not rely on third-party cookies to deliver audiences, and provides ID and audience activations solutions prepared for both now and the future.