CASE STUDY

Customer 360 with AuraDS Results in a 500x Increase in Speed

Startups need scalability, especially when they offer a service that everyone needs. When Orita wanted to build a platform to solve the problem of dirty data, it found a scalable solution in Neo4j AuraDS, a fully managed graph data science offering.

BY THE NUMBERS

500x faster than NetworkX

1.5M clean customer profiles from 3M customer records from 6 sources (sample customer metrics)

PLATFORM

Neo4j AuraDS on Google Cloud

INDUSTRY

Software

USE CASE

Customer 360

OBJECTIVE

Empower e-commerce firms with confidence that their data is clean and complete

CHALLENGE

Bringing order to messy, uneven, overlapping data from multiple sources

SOLUTION

Use Neo4j graph algorithms to create an entity resolution pipeline

RESULTS

- Standardized, validated customer data
- Metrics for customer lifetime value (LTV) and more
- Growing customer base for Orita

The Organization

Orita provides a data confidence platform that helps businesses of all sizes get a better understanding of customers, products, and processes. The platform does identity resolution, making sense of messy, siloed data, using powerful graph algorithms offered in Neo4j Graph Data Science.

The Challenge

Orita Co-Founder Daniel Brady has a data science background as well as hands-on experience at startups. The goal was to start a consulting agency that would empower e-commerce companies to be more data-driven.

The first step was to get some live data to work with. "We did all sorts of projects, from dynamic pricing to getting business intelligence tools online, to helping hire a team of people," said Brady. "But for almost every single client, we built a data warehouse."

But creating that data warehouse required a major chore: bringing order to the chaos of messy, uneven data. In every instance, the clients had numerous data sources and duplicate, erroneous, or incomplete data within those sources.

After using simple methods to unify customer data five or six times, it seemed like a viable product idea. "We keep doing this; why don't we think about a smarter way to do it?" said Brady.

The Strategy

When working with client data, Brady observed that every customer dataset he encountered was "really, really messy." Companies need to know their customers well for many reasons. Customer loyalty, revenue, offers: all of these depend on knowing your customer. But messy data leads to fragmented customer profiles, inaccurate reporting, and more.

"We stumbled upon the world of entity resolution and identity resolution," Brady said. "We decided that there's a real opportunity for us to make a nice standalone product that handles that specifically."

"Our transition from Neo4j Graph Data Science on prem to the as-a-service offer was seamless. With such a small team it is a huge benefit not to have to manage the instance ourselves and do the DevOps and IT management pieces – AuraDS takes care of all of that so we can focus on solving problems for our customers."

Daniel Brady, CEO, Orita

The Solution

The sheer size of the datasets Orita works with meant an identity resolution tool like NetworkX was not tenable – it just took too long and consumed too much memory. But the graph was the correct solution to the problem. The multiplier was Neo4j.

Orita needed a convenient way to store the graph as opposed to loading everything in memory every time. At this point, they looked for something like NetworkX that worked with the Neo4j Graph Database. Orita's product is powered by Neo4j Graph Data Science. The team saw a huge performance leap: "We had about a 500x speed up using the persistent graph database and Neo4j Graph Data Science," said Brady.

From an architecture perspective, Neo4j Graph Data Science is a unified surface, one component that fits seamlessly with the rest of their technology stack. Orita runs graph algorithms on what it calls an "identity resolution pipeline" to unify and clean data for businesses.

Orita switched to Neo4j AuraDS, the fully managed cloud version of Neo4j Graph Data Science. "Our transition from Graph Data Science on prem to the as-a-service offer was seamless," said Brady. "With such a small team it is a huge benefit not to have to manage the instance ourselves and do the DevOps and IT management pieces – AuraDS takes care of all of that so we can focus on solving problems for our customers."

The Results

Orita doesn't stop at identity resolution: it also standardizes and validates customer data, giving customers a clean dataset along with metrics on the dirty data they filtered out. "We go through all the different information from customers and filter out spam emails, fake addresses, and fake phone numbers," said Brady.

Using the remaining clean data, Orita calculates a number of business metrics such as customer lifetime value (LTV), repeat purchase rate, and cohort analysis and provides customers with a detailed report (and the data itself). After Orita runs its special blend of graph algorithms on client data, that cleaning typically means fewer customer entries in the database. But those customers are both real and valuable. "One of the funky things about doing entity resolution is that your LTV can only go up or stay the same," said Brady.

With a mission of turning numerous messy customer data sources into clean, reliable data, Orita is in a growth market. Given that this challenge is nearly universal, the company's prospects are bright.

Questions about Neo4j?
Contact us around the globe:
info@neo4j.com
neo4j.com/contact-us