#### A T S C \Lambda L E

# Rakuten Seamlessly Migrates 10 Million Customers from On-Premise Data Warehouse to Snowflake with AtScale

Eliminates computing resource limitations and business disruption

Thanks to AtScale, adopting Snowflake was the easiest part of our cloud transformation. We just repointed AtScale from our old Hadoop environment to Snowflake, and it was seamless for the users.

> - Mark Stange-Tregear Director of Analytics, Rakuten

## **Executive Summary**



Rakuten connects savvy shoppers with America's best brands to give them a hassle-free way to save money on the things they buy every day. Founded in 1999, Rakuten now has over 10 million users, has helped members earn over \$1 billion in cash back, and is one of the largest free-membership loyalty programs in the U.S.

#### The Challenge

Rakuten invested in a data-driven business environment and customer experience by consolidating their data into an on-premise data lake using Hadoop coupled with AtScale's virtual OLAP cubes to analyze their data. After initial success, the Hadoop cluster's physical limitations began to cause painful business disruptions as disk and processor usage hit critical levels. Rakuten needed a new data warehousing solution that would ensure responsive data analysis during peak demand periods.

Á.

#### ATSCALE

#### The Solution

Rakuten would restructure their data infrastructure by moving from their on-premise Hadoop cluster to a Snowflake cloud data warehouse on AWS. Colder, less frequently used data would reside in S3 on AWS, while hotter, high-demand data would reside on Snowflake. AtScale would provide query optimization and semantic connections to the data.

#### The Results

Snowflake's elastic and scalable resource model allows Rakuten to bring additional computing power online during peak demand periods to maintain responsiveness to queries. AtScale provides labor-saving automation for Rakuten: queries are easier to build, aggregates are built automatically, and since AtScale acts as an abstraction layer for Rakuten's BI tools, the switch from Hadoop to Snowflake was completed without any interruption of data service.

#### **Benefits at a Glance:**

- Provide surge capacity during peak data usage
- Seamless transition to the cloud without disruption
- Optimize query response times and conserve resources
- Easily explore the data warehouse for insights
- Avoid hiring additional data scientists

# Rakuten's Cloud Transformation Deep Dive

## A Data-Driven Vision

Rakuten's vision is to arm all of their employees with the ability to derive insights from their company data. The company leverages member data such as who shops where, when, and how often, as well as prices, commissions, and more, to craft compelling offerings and satisfying service. To accomplish this, they offer employees access to their data warehouse through Tableau Server, drag and drop data analysis via Tableau Desktop, and even the freedom to write their own SQL queries.

#### ATSCALE

#### Why Finite Resources Required a New Approach to Data Warehousing

Rakuten's data warehouse had been consolidated (out of many siloed systems) onto a single, on-premise Hadoop cluster. At first, the solution was a huge success, providing levels of fast, rich data access that were previously impossible. But as the system became more widely used within the company, the limitations of an on-premise solution began to show.

Rakuten discovered that the resources of an on-premise computing cluster were finite, with troubling consequences. Competition for hard disk access, memory, and CPU time from business units or resource-intensive processes would slow down or even freeze the cluster. For example, the finance department's activity surges at the beginning and end of each month, causing spikes in demand over short, recurring intervals. If the marketing department ran a large-scale email campaign near the beginning of the month that required a lot of compute, finance would have a hard time running their month-end close calculations from the previous month.

Other times, queries would drain massive resources from the Hadoop cluster. A user once rewrote the script for a detailed marketing report, increasing the amount of data analyzed from two weeks to two years. The runaway query resulted in unnecessary joins on tables containing billions of rows, resulting in the data cluster effectively freezing up as it ran out of resources. Rakuten found the offending query and fixed it to prevent any recurrence, but since Rakuten runs 1.5 million queries against their data warehouse each month, they realized this was not a scalable strategy.

#### The New Performance Solution

In 2018, Rakuten migrated its high-demand data to Snowflake, with AtScale providing query performance optimization and providing a single, virtualized view of their data delivered as a Data-as-a-Service.

Snowflake on AWS provides flexible compute resource allocation and resource usage statistics:

- Their architecture separates storage from computing resources, enabling Rakuten to scale up or down as needed to avoid slowdowns in query response times.
- 2. Snowflake can create accounts for each of the enterprise's departments and

#### A T S C \Lambda L E

business units, allowing Rakuten to track which departments are using the most data warehouse resources. Rakuten can use this data to identify data hungry departments in need of more support, or groups that might be running inefficient queries.

#### AtScale streamlines Rakuten's data on Snowflake in two ways:

- AtScale make queries incredibly responsive. A huge number of data requests require only a few hundred data points out of the 50 thousand data points stored in the Rakuten data warehouse. AtScale's Intelligent Aggregates use only the data needed, shortening response times and economize computing resources.
- 2. AtScale's design tool allows business users and data scientists to quickly build queries that allow them to explore the data within Rakuten's data warehouse without the need to create complicated SQL scripts.

Without AtScale, query performance is too slow. We would have to devote significant time to building aggregates by hand to even approach the efficiency and speed AtScale provides. AtScale's ability to automatically create and manage highly-efficient aggregates is critical to our team's success.

## Switching to the Cloud

— Mark Stange-Tregear Director of Analytics, Rakuten

#### Two things happened the day Rakuten switched on their cloud data warehouse with AtScale:

- **1.** There was an immediate 30% drop in load on the on-premise computing cluster as high-demand queries shifted toward Snowflake and AtScale.
- 2. None of the business users noticed the switch. There was no disruption for business users whether they were using Excel, Tableau, or SQL.

Rakuten eliminated business disruptions, ensured speedy query response times, and secured ample room for their data warehouse to grow with their business by migrating to the cloud with Snowflake and AtScale.



# ATSCALE

