Sharethrough Streamlines the Analytics Pipeline for Greater Performance with Snowflake

sharethrough

snowflake | **Fivetran** | **MicroStrategy**   CASE STUDY

Sharethrough is the largest independent native advertising platform, used by leading global publishers like Time Inc., Wenner Media, Scripps Networks Interactive, Thought Catalog, Warner Bros. and USA Today Sports to manage all aspects of native ad monetization. More than three billion native ad impressions are served every month through the Sharethrough Exchange into the feeds of these top publishers.

Native ads are read, not just seen like banner ads, delivering meaningful and earned audience attention that allows brands to create positive brand associations and influence brand lift. Major brand advertisers, including 49 of the 50 largest brands in the world (Microsoft, Intel, ConAgra and hundreds of other household names), use Sharethrough's native ad-buying platform Sharethrough for Advertisers (SFA) to promote their content on publisher sites.

Sharethrough's native ads fit the form and function of the surrounding editorial wherever they're placed. Understanding the user behavior and experience of the audiences that are engaging with its ads is central to its business model. Its analytics focus on how well websites are performing, how content renders and what can be done to improve the viewing experience for its ads.

> "We went from using 14 technologies to a stack of four powerhouses— AWS, Snowflake, Fivetran and MicroStrategy—that can be managed by a single person."
>
> — *Joseph Bates, Head of Analytics, Sharethrough*

THE CHALLENGE

Complicated Pipeline, Limited Analytics Success

Sharethough had outgrown the capacity of their data analytics environment which was custom-built using a MySQL product. With no semantic layer, poor performance and a maxed-out row capacity, getting data for analysis was an extremely complex and manual operation.
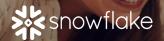
Pulling data from 2 databases, one for metadata and one for fact data, regular run-of-the-mill reports took half an hour. Sharethrough was forced to create multiple new tables at specific granularities. As the tables grew and the schema became more complex, it reached a point where most every user had to understand the underlying schema of the data warehouse. "You really don't want your business users thinking about how the data is physically architected." exclaimed Joseph Bates, Head of Analytics.

THE SOLUTION

Tap into Raw Data for Rapid Analysis

One of the first changes that Sharethrough made was to remove the dependency for MySQL. By bringing in Snowflake as the core warehouse, Sharethrough was able to tap directly into the raw data stream from JSON sources. Snowflake enabled Sharethrough to take all website activities such as raw JSON logs directly from Amazon S3 and ingest them directly into Snowflake without preprocessing. This allowed Sharethrough to bypass Spark and directly access S3 to obtain the logs.

Using Fivetran, Sharethrough further enhanced the extraction and load process by making it easy to load different data sources. For example, Sharethrough could now move data from its SaaS applications, marketing applications and SalesForce.com easily to Snowflake for rapid transformation to create the views required by analysts. Today, Sharethrough connects four different sources through Fivetran, and all of them sync in less than two hours. In total, Sharethrough now runs 40 terabytes of mixed data through Snowflake and reduced costs by eliminating the need to store data in temporary locations. "The increased direct access to data with Fivetran and Snowflake made a huge difference in our productivity," Bates said.

High Performance Analytics with Less Resources

"The Snowflake and Fivetran interfaces are incredibly easy to use," Bates said. Inside Snowflake, a single click on a JSON column pulls up the data in a smartly parsed and formatted way that is ready to use. "In seconds, I can accommodate a user request for a new field in the data stream, and there is no need for any architectural work."

Implementation of Snowflake and Fivetran proved very simple. With only one-full-time staff person, Sharethrough was able to stand up an entirely new data warehouse of 40 terabytes of mixed data within weeks and connect it to data sources. Additionally, no special skills for database management and teams of consultants and engineers were necessary. "With the manpower that we had, Snowflake and Fivetran were the right tools for the job and we couldn't be happier with the performance," Bates said.
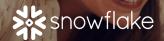
Make Complex Data Available to Business Users

While the engineers and product teams use Snowflake directly, Sharethrough brought in MicroStrategy to make the data available to the nontechnical user. MicroStrategy is extremely sophisticated in how it generates queries through to Snowflake. With the new technology stack, Sharethrough sends new test and production data immediately to business users through MicroStrategy for rapid analysis and decision-making. "What used to take one month, now takes hours," Bates said.

Sharethrough runs 500 to 1,200 reports a day with an average creation time of less than four seconds and average query size of 700 megabytes. The team now has a dynamic dashboard that models their own visualizations. With the added data types and a semantic layer in place, the complexity of the queries allows for more powerful in-depth analysis.

"With strong support for the Snowflake and Fivetran infrastructure, being able to lay a really sophisticated, powerful product like MicroStrategy on top of the whole thing is super cool—and that's where I think our technology stack is unique." Bates said. Sharethrough is considering making the entire solution available to its end users—30,000 or more ad buyers in total.

RESULTS

High Power in a Short Stack

Typically, queries that would be saved up and run over a weekend, now run in 30 seconds and take four lines in SQL. With this short stack, Sharethrough performs analytics 2,000 times faster. "I'm so pleased with the performance for the more complex jobs with higher granular data and jobs which run aggregations, de-duplication and frequency-level analysis," said Bates. "Reports that were taking 45 minutes to run in MySQL take half a second in Snowflake." Sharethrough went from an analytics stack 14 technologies deep, to just four (with AWS, Snowflake Computing, Fivetran and MicroStrategy).

RESULTS SUMMARY

- Achieved 2,000 times performance increase
- Reduced analytics stack from 14 technologies to four
- Simplified environmental management to a single staff member
- Provide infinitely scalable data warehouse
- Transformed all technologies to cloud service for lower-cost scalability

ABOUT SNOWFLAKE

Snowflake is the only data warehouse built for the cloud. Snowflake delivers the performance, concurrency and simplicity needed to store and analyze all of an organization's data in one location. Snowflake's technology combines the power of data warehousing, the flexibility of big data platforms and the elasticity of the cloud at a fraction of the cost of traditional solutions. Snowflake: Your data, no limits. Find out more at **snowflake.net**.