Jana uses Snowflake to Provide Faster Data Insights Across the Company

JANA

CUSTOMER Jana

PARTNER Snowflake

snowflake

CASE STUDY

Jana provides free, unrestricted internet access to more than 30 million smartphone users in emerging markets. With their mCent Android app, Jana shifts the cost of mobile data to brands via sponsored content. When users engage with content in mCent, they earn free mobile data that can be used anywhere on the internet. Jana product managers use data to constantly analyze how well each of the features in mCent performs, to determine which features should be turned on for everyone, and which features should be turned off.

THE PROBLEM

Creating Compelling Internet Products

To date, mCent has featured more than 4,000 brands in over 15 emerging markets. Jana collects and analyzes mCent user behavior data to understand the best ways to encourage users to interact with those brands in each market. When new branded content or mCent features are introduced, Jana uses split-tests to analyze and measure key metrics, such as user attention, lifetime value of the user, and KPIs that are important to brands.

In addition to understanding user behavior, Jana works to address fraudulent behavior within mCent. At Jana, both the sales and marketing teams access data to drive new business and create collateral and insights that enable their partners to understand the value of mCent. In addition, Jana has a series of systems that monitor the latency of business constraints and catch bugs, identify fraud, and recognize unexpected developments in the various countries.

Already driving more than 4.0 billion MB of app usage, Jana needed a scalable and cost-effective solution to process and analyze all its data.

THE CHALLENGE

Outgrowing MySQL

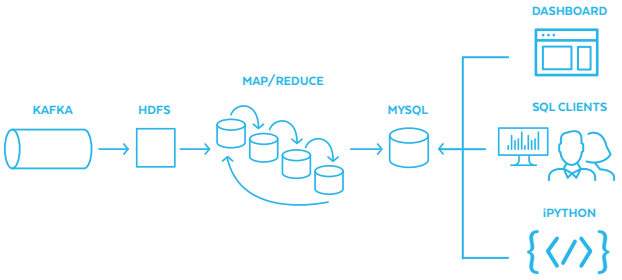
Jana had set up its initial architecture for the analytics system while the company was an early-stage startup, before there was a large amount of usage.

"Our key goal was to get something together quickly that would work for the time being," said Craig Lancaster, CTO of Jana. "The structure could remain for the long run, but we knew that we might have difficulties scaling certain parts of it."

In the original system, all code was deployed in Amazon Web Services using MySQL. Web applications and the server-side API for the Android application generated the bulk of the data. The code was then written in Python using the Flask web framework behind Nginx and stored within Cassandra. Jana analyzed 37 distinct types of data, including event counting, accounting records, user profiles, and campaign configuration. All data was encoded with Avro.

"Initially it worked very well," said Lancaster. "But challenges arose as our business scaled."

ORIGINAL ARCHITECTURE



Because much of the infrastructure had remained static since the initial build, data outgrew what MySQL could handle. The expanding Jana team also needed a more efficient table design: as the system scaled, queries became slower and slower and table scans became unfeasible. More and more administration time was required to add capacity and backup systems and to administer the Hadoop cluster.

"We needed to redesign our system before we were unable to run the business," said Lancaster.

WHY SNOWFLAKE?

Finding a Better Approach

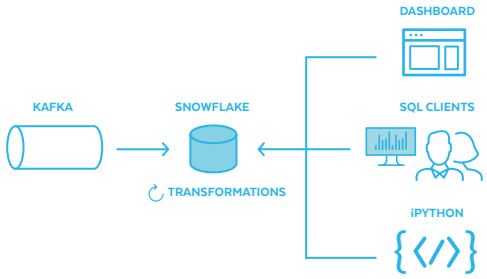
For Jana, the growing usage of data was reaching a point where they could no longer operate the business efficiently. Because Jana was mostly happy with its original architectural approach, the challenge was to figure out how to upgrade the parts of the system that were becoming bottlenecks, while retaining the other benefits. In addition, Jana hoped to reduce its growing administration and overhead costs.

While the overall structure of the analytics system remained unchanged, Jana upgraded most of the key components of the data pipeline. For data processing, Jana replaced MySQL with the Snowflake Elastic Data Warehouse.

"The fact that 80% of our employees actually access the database is astounding to me. I think that is a testament to how easy it is to access Snowflake."

— Craig Lancaster, CTO, Jana

CURRENT ARCHITECTURE



"We simply copied the files we had been loading into MySQL into S3 and then issued a copy command in Snowflake," said Lancaster. "Immediately we began to run queries and analyses that we had become unable to finish in MySQL. We backloaded months of data into Snowflake in less processing time than it took to load one day's data into MySQL."

"With Snowflake's semi-structured data handling, we have replaced a series of map-reduce steps with three or four SQL statements that we're directly accessing in the Avro data and loading into the same structured tables that our analysts were accustomed to using," said Lancaster. "We're also able to issue some of the queries directly on the Avro data without having to de-structure them, because the performance was so much improved due to the indexing and optimization within Snowflake."

Jana leveraged Snowflake to keep up with the business demands of processing and analyzing a rapidly growing stream of data. Jana achieved a number of benefits, including:

- Increased use by analysts
- Scalability for improved testing and faster time to market for new features
- Significant reduction in administration overhead

THE RESULTS

Increased Use By Analysts

"The better performance that we see with Snowflake, and the simplified table structures that we put into place as we migrated, has increased usage within the organization," said Lancaster.

Due to the introduction of Snowflake, the user base of the environment has doubled.

"The fact that 80% of Jana employees actually access the database is astounding to me," said Lancaster. "I think that is a testament to how easy it is to access Snowflake. In particular, the web interface lowers the barrier to entry and enables us to get people into the system who have never issued SQL queries in the past at any previous job, enabling them to be productive very quickly."

Scalability For Improved Testing And Faster Time To Market For New Features

"Snowflake makes it easy for us to add additional warehouses for urgent analysis," said Lancaster.

During migration to Snowflake, Jana leveraged the performance of the cloud data warehouse to improve testing in the data pipeline. By running different warehouses against existing data, Jana experienced complete isolation between workloads.

"It's trivial for Snowflake to spin up separate warehouses to test against, which allows us to run testing workloads without affecting the production workloads," said Lancaster. "The ability to test against production data without

"Immediately we began to run queries and analyses that we had become unable to finish in MySQL. We backloaded months of data into Snowflake in less processing time than it took to load one day's data into MySQL."

— Craig Lancaster, CTO, Jana

interfering with existing workloads helps in bringing new features to market quickly."

Significant Reduction In Administration Overhead

The biggest benefit from the engineering side has been significantly reduced administration overhead. Snowflake makes available compute resources with a couple of clicks on the interface, which has freed up time.

"Snowflake, and Amazon S3 are each trivially elastic," said Lancaster. "Capacity planning is simply something we don't have to worry about for any of those any more."

In addition, Jana has reduced the time required to tune table designs or database parameters.

"Snowflake's logical design and well-written SQL queries give us the performance we need," said Lancaster. "As an added bonus, we don't need to worry about backups or disaster recovery. With Snowflake, our operations team regained 25% of their time compared to the previous system."

Concentrating On Business Advantage Rather Than Operations

Using the Snowflake Elastic Data Warehouse, Jana processes and analyzes app usage data in a highperformance, scalable way without the cost and complexity of other solutions.

"The net effect with Snowflake is that we're able to focus more on our products and on better understanding our business, and less on simply trying to operate our analytics system," said Lancaster.

> "With Snowflake, our operations team regained 25% of their time compared to the previous system."
>
> *— Craig Lancaster, CTO, Jana*

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**.