SAINSBURY'S ENGAGES WITH PROFESSIONAL SERVICES TO OPTIMIZE SNOWFLAKE AND CREATE CENTER OF EXCELLENCE

RETAIL/CPG

Sainsbury's

COMPANY	Sainsbury's
LOCATION	London, England

KEY METRICS

72%
Reduction in job length

25 times
Increase in throughput per minute

SAINSBURY'S TRANSFORMATION WITH SNOWFLAKE

Sainsbury's, the U.K.'s second largest retailer with 2,200 stores and a large digital presence, offers customers a wide range of distinctive, quality products at competitive prices. Recently, its leadership transformed the company, merging all their different offerings into one multi-brand, multi-channel business, eliminating data silos, and making data available to everyone in the company. They chose Snowflake on AWS to be the cornerstone of their transformation and retained a Resident Solutions Architect (RSA) through Snowflake's Professional Services team to help bring their goals into reality.

CODESIGNING A SIMPLER ARCHITECTURE

On average, Sainsbury's Online Grocery ingests 30 to 40 million events per day in real time, and holiday peaks easily double the number of events. The Sainsbury's team discovered that Snowflake's Streams and Tasks feature could enable them to scale up and down to manage fluctuating data loads. They consulted Hilda Davies, their Snowflake RSA, for help with optimization. "We do a lot of heavy lifting in Snowflake to build our products. Making something work isn't difficult to achieve. But achieving the best possible system takes collaboration and learning best practices. Our efforts have paid off, and we're doing things we previously didn't think were possible," said Sainsbury's Data Engineering Manager, Joan Fuerte. "It's worth our time to become Snowflake experts."

With Streams and Tasks, Fuerte's team and Davies streamlined and simplified the architecture, as shown in Figures 1 and 2. According to Davies, "The simplified overall design is easier for people to understand, and as a result we got better performance and reduced costs."

> **I'm aiming to deliver fine-tuned, really efficient, and good products. I need to rely on all the options I have, and I cannot stress enough how important and great the work our Snowflake RSA has done with us."**
>
> **—JOAN FUERTE,** Data Engineering Manager, Sainsbury's
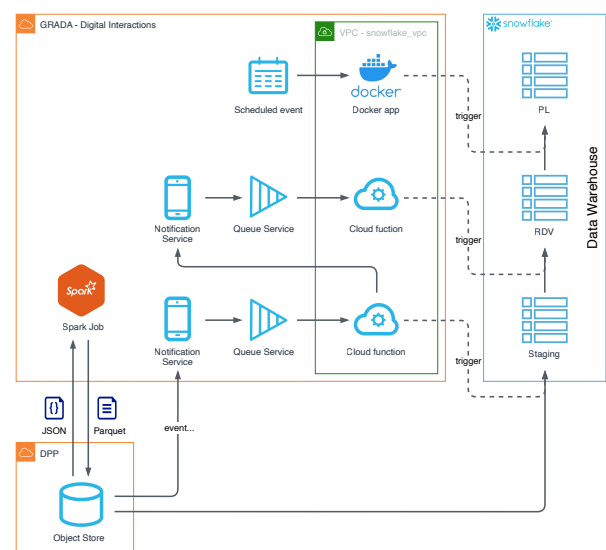

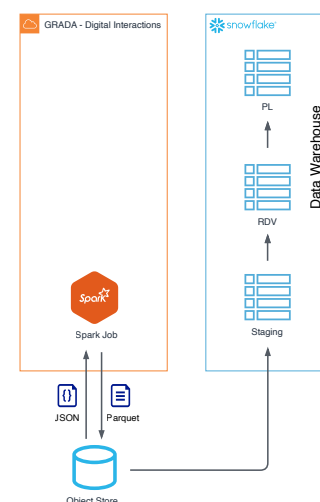**Figure 1: Managing fluctuating loads with the original architecture**

Figure 2: Managing fluctuating loads with a simplified architecture

OPTIMIZING SNOWFLAKE WITH BEST PRACTICES

In addition to creating a simplified architecture, Fuerte's team wanted to understand how to determine the optimal data warehouse size for their varying ETL workflows. Until that point, everything had been running in a one-sized warehouse.

After consulting Davies, they created monitoring dashboards. When they reviewed the dashboards, they realized they could reduce their 2XL warehouse size significantly. According to Jody Buczynski, Sainsbury's Data Engineer, "We had a pipeline that was running nearly all the time. We moved it down to a medium sized data warehouse, and that one change instantly reduced our cost. This is incredibly powerful, not only to save money, but being able to monitor that this is a safe procedure and won't impact performance."

With the old pipeline, scheduled batch imports of Amazon S3 files into Snowflake (using a combination of AWS infrastructure and SQL code) took two to three sprints (four to six weeks) to run. With the new pipeline's simplified architecture leveraging Snowflake streams and tasks, ingesting and converting staged Amazon S3 files takes two to three days, with easy-to-follow patterns and less debugging. Figure 3 shows a 72% reduction in average minutes per job execution and more than a 25x increase in average bytes written per minute.

Sample Date	March 29, 2021 from 9:00 to 12:00	
	OLD PIPELINE	NEW PIPELINE
Average Minutes Per Job Execution	18	5
Minimum Minutes Per Job Execution	11	4
Maximum Minutes Per Job Execution	34	5
Average Bytes Written Per Minute	7,968	210,232
Average Bytes Written Per Job Execution	255,121,920	67,923,304
Minimum Bytes Written Per Job Execution	160,106,496	50,907,750
Maximum Bytes Written Per Job Execution	460,641,280	81,846,541
Average Rows Inserted Per Job Execution	145,075	210,232
Minimum Rows Inserted Per Job Execution	92,723	129,217
Maximum Rows Inserted Per Job Execution	267,936	253,232

Figure 3: Pipeline efficiency improvements
Reduced average throughput shows greater efficiency, and reduced variation between min and max throughput shows consistent performance, making predictability and provisioning easier.

CONTRIBUTING TO SAINSBURY'S CENTER OF EXCELLENCE

"Snowflake as a technology is already great, but we've gained a massive amount of value to maximize its abilities through our conversations with Hilda," according to Sainsbury Data Engineer Jody Buczynski. After learning from Davies, the Sainsbury's team gained expertise that enabled them to be more self-sufficient. "We can now make decisions about other key parts of the pipeline with much more confidence in what we're doing, and to me, that's really amazing," Buczynski said.

Fuerte's team actively shares their learnings and expertise with the rest of Sainsbury's. For example, anyone can use their solution to optimize data warehouse sizing. "There was another team running on a 2XL data warehouse. It would take 25 minutes to query a whole table, and they were running multiple queries a day. That's a massive cost," said Buczynski. "We applied our own learnings and helped them resize to a large data warehouse, and queries now complete in less than a minute. The cost difference is just phenomenal."

> I really enjoy conversations with Hilda—it's a special relationship. Don't get me wrong, we could just ask her questions, and she can tell us what to do. But instead we seek her advice, and she advises well. She is the Snowflake expert and we are the Sainsbury's experts. To achieve our end goals, we need to collaborate from the beginning to get the best possible results."
>
> —JODY BUCZYNSKI, Data Engineer, Sainsbury's

CONTINUING THE PARTNERSHIP WITH SNOWFLAKE PROFESSIONAL SERVICES

"We really view our Snowflake RSA as an extension of our Sainsbury's team," said Fuerte. "We are now in a position to deliver new use cases at a higher velocity." Fuerte's team now focuses their energy on important business projects and nurturing Snowflake expertise among other Sainsbury's teams. "Our roadmap with Snowflake is long, and we're excited to continue the partnership and create impactful applications for Sainsbury's," according to Fuerte.

> The team is very open with me. I get the holistic view of what they're trying to accomplish so we can collaborate effectively from the very beginning."
>
> —HILDA DAVIES, Resident Solutions Architect for Sainsbury's, Snowflake

ABOUT SNOWFLAKE

Snowflake delivers the Data Cloud—a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud. **snowflake.com**