QWILT with yuki

How Qwilt **Cut 63%** of Snowflake Costs While Improving Performance

QWILT

About Qwilt

Qwilt is a modern CDN powering streaming, gaming, and massive downloads for some of the world's largest content platforms.

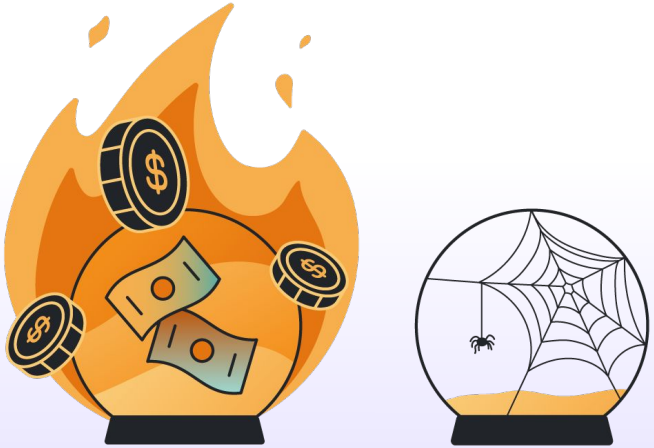
Their data team built a mission-critical **customer-facing data application on Snowflake**, serving dozens of BI requests per second aimed at analyzing traffic patterns of millions of requests per second that were handled by the Qwilt network.

QWILT

At peak, Qwilt handles over 6 million requests per second, processing high-scale data across Snowflake. Their stack includes Snowpipe Streaming, Kubernetes autoscaling, and tight SLAs for real-time data access.

Software Development

Challenges



Mission-critical data app on Snowflake

Strict SLAs for BI queries powering customer-facing analytics.

Unpredictable Snowflake costs

Even after months of tuning, query compute remained volatile, making spend hard to forecast and keep under control.

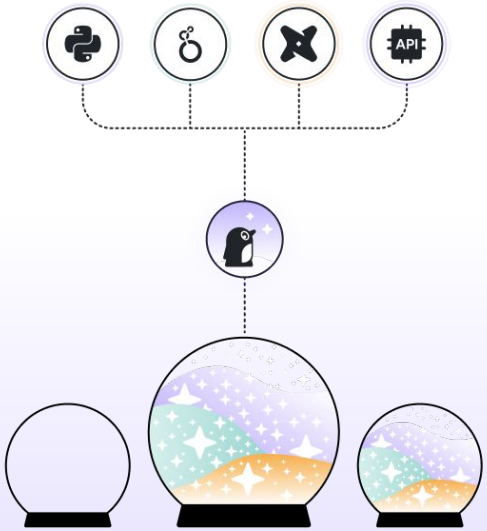
Manual effort, limited scalability

The team invested heavily in internal cost governance logic, but as the platform grew, engineering effort alone couldn't keep pace.

Real-time workloads under pressure

Multiple user personas, time zones, and real-time SLAs made warehouse right-sizing a challenge.

Solution



Dynamic warehouse optimization

Query metadata analyzed in real time, keeping customer-facing requests fast while routing workloads to the most efficient warehouse.

Fully private deployment:

Runs inside Qwilt's environment - No data leaves their cloud.

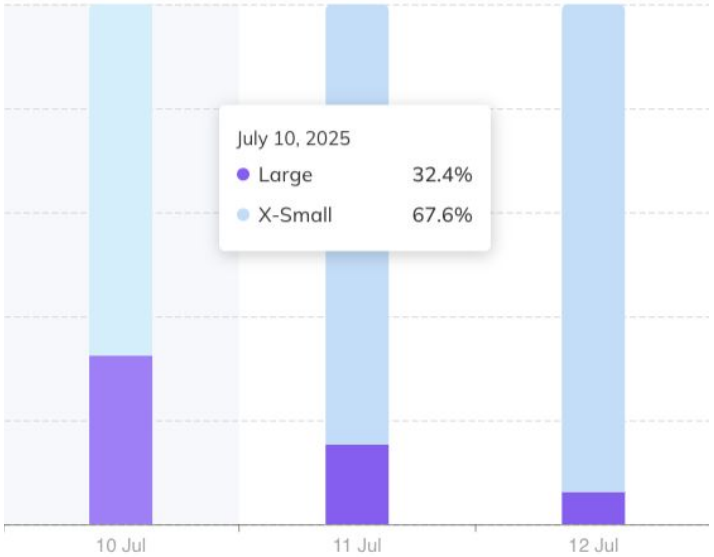
Tenant-level insights

Granular Clear visibility into which workloads truly required larger warehouses and more compute.

Results

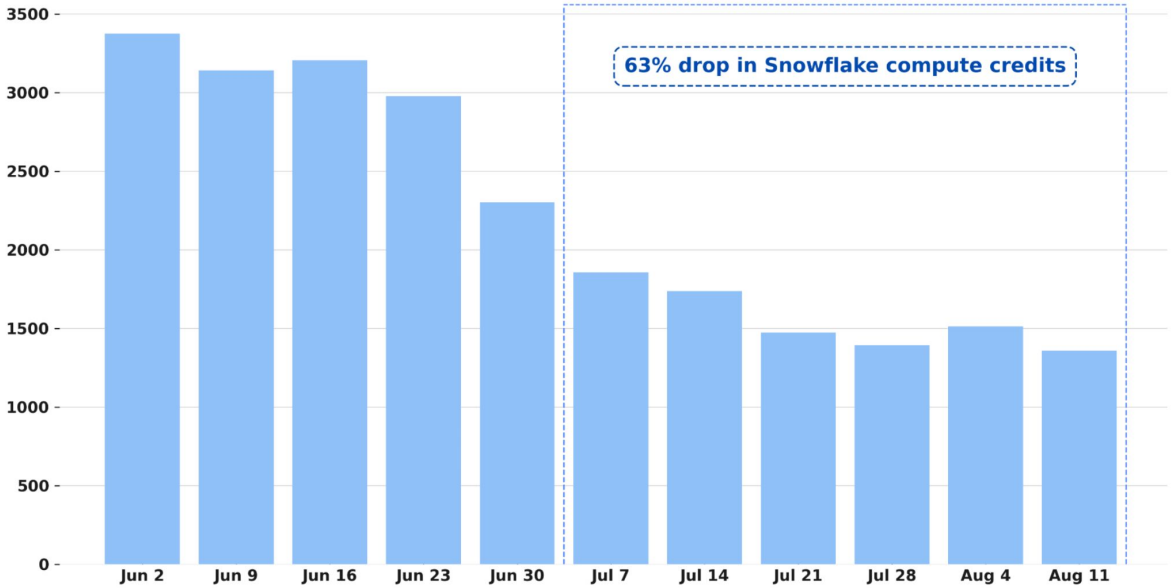
The impact was immediate:

- Only 32.4% of queries needed a large warehouse
- 67.6% ran just as fast on an X-Small



By isolating heavy queries from lighter ones, Yuki eliminated resource contention, reduced queuing, and improved responsiveness across workloads.
Everything was re-routed automatically without any manual intervention.

Results



63%

Cost Reduction

Within just a few days.

Snowflake compute costs dropped and remained consistently lower.

With Yuki handling warehouse optimization automatically, **costs stabilized at a lower baseline** and engineering hours were freed to focus on higher-impact projects.

End-users experienced consistent performance even as costs dropped.

> We've spent months building internal rules to manage Snowflake spend. Query filters, time slicing, and cost controls across users. It helped, but it was a constant effort. When we plugged in Yuki, I honestly didn't expect much. Vendors always say 'one-day integration,' but this time, it actually delivered. Within 24 hours, we saw a 63% drop in compute costs. It just worked.

Ron Kitay,
Software and Data Engineer at Qwilt

yuki

> Working with Ron and the Qwilt team was the perfect stress test for Yuki. They had already done the hard work: internal rules, smart query shaping, deep cost analysis. And still, Yuki was able to take them further. Seeing 40% of queries auto-right-sized in week one and a 63% cost reduction without touching a line of code. That's exactly why we built this product.

Amir Peres,
CTO at Yuki

Get a free analysis!

Find out how much you can save

Free Analysis