Browsi with yuki

# How Browsi Scaled **Customer-Facing AI Agents** on Snowflake with Yuki

Browsi

# About Browsi

Browsi powers one of the industry's largest real-time advertising intelligence platforms, helping brands and agencies understand how competitors are spending and scaling campaigns.

With its AI-driven product, Browsi turns billions of advertising signals into a **conversational intelligence experience** letting customers ask complex questions and receive
actionable insights instantly.

Browsi

**Data Stack**

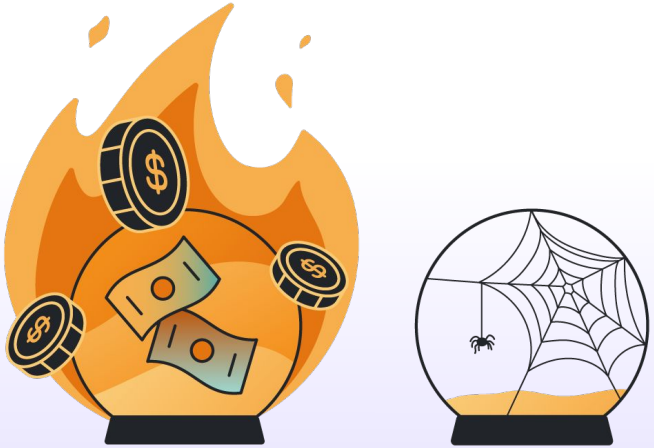
snowflake

dbt

AI Agents

**Real-Time Ad Intelligence Platform**

# Challenges



**SLA at risk during peak traffic**

Every customer request needed to return within a set time window. Browsi couldn't confidently guarantee this under real traffic, making production rollout risky.

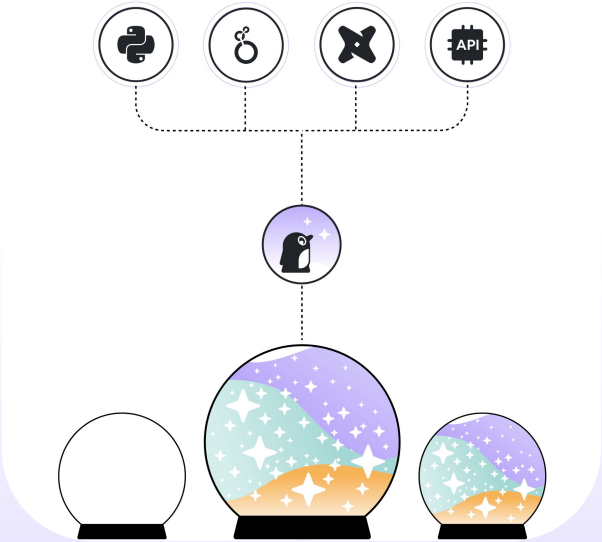
**Forced to overprovision compute**

To protect response times, the safest path was running larger warehouses than needed most of the time - a costly trade-off that was unsustainable at scale.

**Unpredictable query workloads**

Some customer questions triggered lightweight queries. Others kicked off heavy analysis. When multiple customers hit the AI agent simultaneously, queue times spiked without warning.

# Solution




### Responsive during traffic spikes

Customer-facing traffic stays responsive during spikes without keeping a large warehouse running all day “just in case.” Latency-sensitive requests get priority when it matters.

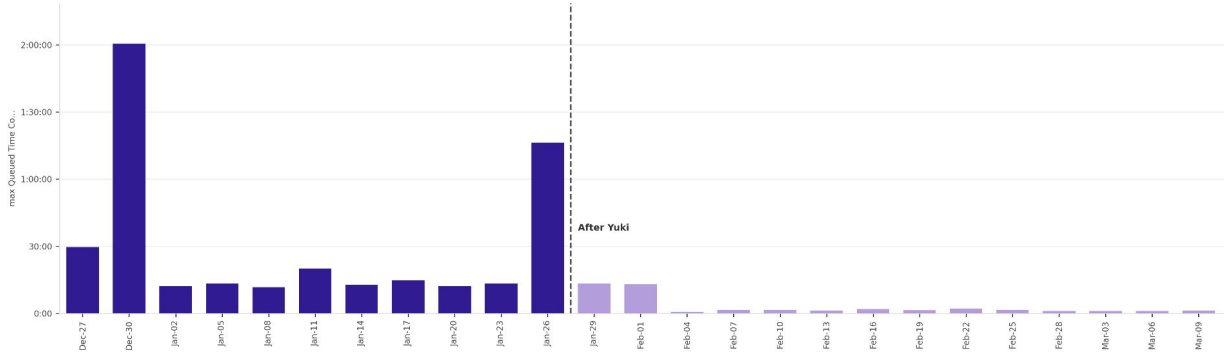
### Intelligent per-query routing

Yuki sits in the path of every query the AI agent generates, automatically routing it to the right warehouse so lightweight requests don’t clog compute that heavier queries need.

### Zero-touch setup, full visibility

No query rewrites, no application layer changes. From day one, the team could see where credits were going and how queries were being routed.

# Results




# 81%

## Avg Queue Time Reduction

**Within 24 hours of go-live.**

Average queue time dropped **from 12.9s to 2.5s per query.**

Peak queue time fell **89%** - from over **2 hours down to 13 minutes.**

The numbers that had made production go-live impossible were gone. Browsi could now scale their AI agent with **confidence and cost efficiency** - without adding operational overhead.

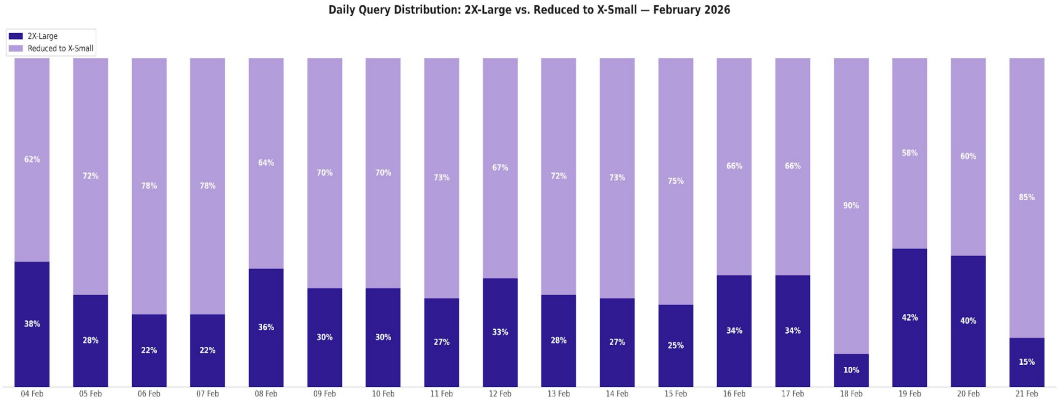
# Results

## 28% credit reduction

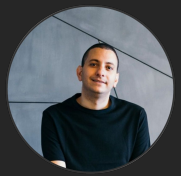
Within the first month, Browsi reduced Snowflake credit consumption by 28%. 32% of queries were automatically routed to a smaller warehouse - no manual tuning required.

## 2.5x workload, same cost

Monthly query volume more than doubled - from 21K in January to 55K in February - while credits consumed stayed roughly flat. **More than 2.5x the workload at nearly the same cost.**



> Yuki helped us turn our AI agent into something we could scale quickly with high cost efficiency. We protected the customer experience during spikes and stayed disciplined on spend, without adding operational overhead.



Asaf Shamly

Co-Founder & CEO @ Browsi

yuki

> Most companies try to solve AI-agent traffic by buying bigger warehouses. Browsi did it the smarter way: they kept performance predictable by controlling routing, not by overprovisioning. That's exactly what Yuki was built for.

Ido Arieli Noga,
CEO & Co-Founder @ Yuki

# Get a free analysis!

Find out how much you can save

Free Analysis