

TRANSFORMING REAL-TIME DATA AT SCALE:

How Wunderkind Reduced Costs by 20% with Deephaven

Written by: Christian Saide
Principal Software Engineer at Wunderkind

CONTENTS

■ Wunderkind Background	01
■ The Challenge	02
■ The Problem	03
■ The Requirements	04
■ The Solution	05
■ Before + After Architecture	06
■ The Results	07

WUNDERKIND BACKGROUND

9 BILLION DEVICES ARE RECOGNIZED

Wunderkind leads in AI-driven marketing, transforming brand-audience connections with hyper-personalized messaging. Our Autonomous Marketing Platform leverages first-party data, identifying anonymous traffic via a network that recognizes 9 billion devices. We enable targeted offers at the perfect moment across email, text, and ads, integrating with brands' systems for seamless performance enhancement.

\$5 BILLION IN ANNUAL REVENUE

Delivering over \$5 billion annually in attributable revenue, we guarantee results, making us a top revenue channel for clients like Harley-Davidson and Shoe Carnival. Choose Wunderkind for unmatched precision in boosting your top-line revenue.

THE CHALLENGE

In their own words:

Wunderkind operates across thousands of websites and applications, collecting information from end users on behalf of our clients. This staggering amount of data ranges into the trillions of rows across pebibytes.

THIS DATA IS COLLECTED AT RATES REACHING HUNDREDS OF THOUSANDS OF EVENTS PER SECOND,

and all of it must be analyzed and acted upon in real time.

This analysis powers the engines behind our decision-making, which in turn activates target marketing campaigns for our clients. This allows for highly granular targeting and personalization of content and promotions.

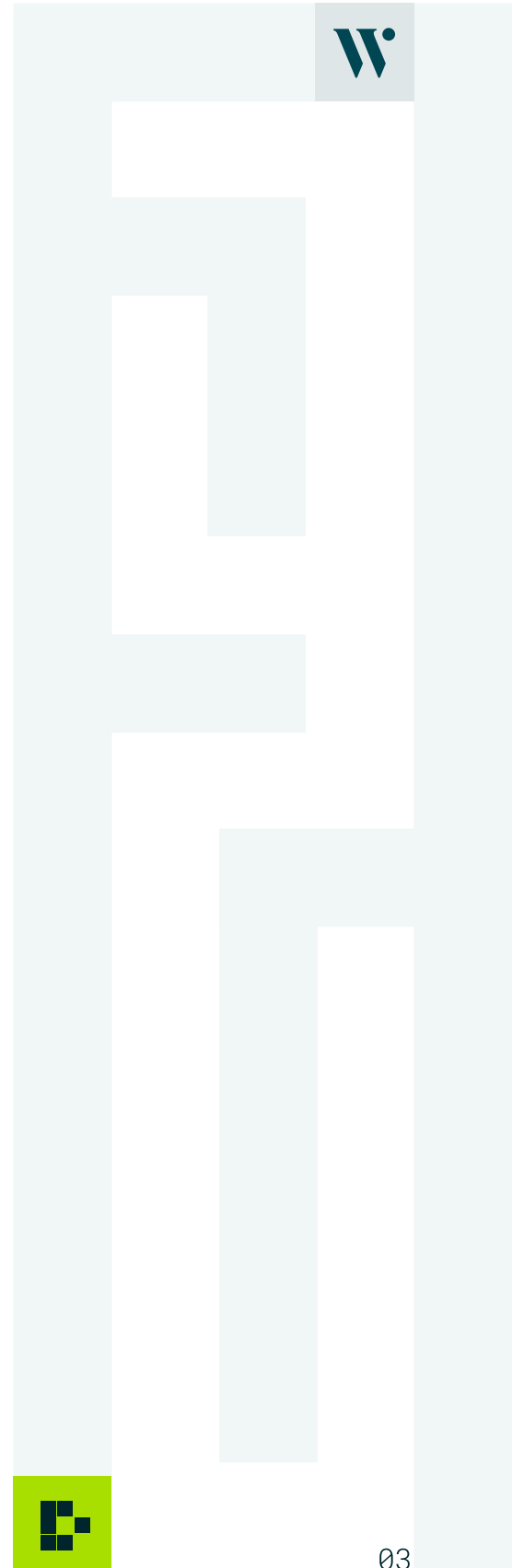
THE PROBLEM

We are a highly talented but small team that needs to quickly and easily manage our data and workflows. We also require tools that can scale to meet the demand without requiring huge time and resource investments.

The existing tool landscape left us with either bespoke custom implementations to handle our use cases or one of the incumbent ETL solutions, such as Spark or Dataflow.

The underlying problem we faced spanned a few different dimensions:

- Existing technologies require specialists to operate at scale, of which we had very few on the team.
- Due to our somewhat unique requirements, the existing ETL solutions didn't work well out of the box.



THE REQUIREMENTS

Our requirements were simple:

Support for analyzing/ aggregating hundreds of thousands to millions of messages a second.

An efficient and simple development process that any backend engineer can slot into.

Support for E2E latencies of sub-10s from the moment of collection to being actionable in our algorithms.

A flexible toolbox of functionality we can use to build out our pipelines.

Out-of-the-box composability of pipelines.

A rich community for support and ideating.

The last two requirements were key to our success. We knew we needed a setup that would allow us to break out of the existing ETL pipeline and create custom implementations when needed. Additionally, we sought an active community with strong support that we could lean on as needed.

THE SOLUTION

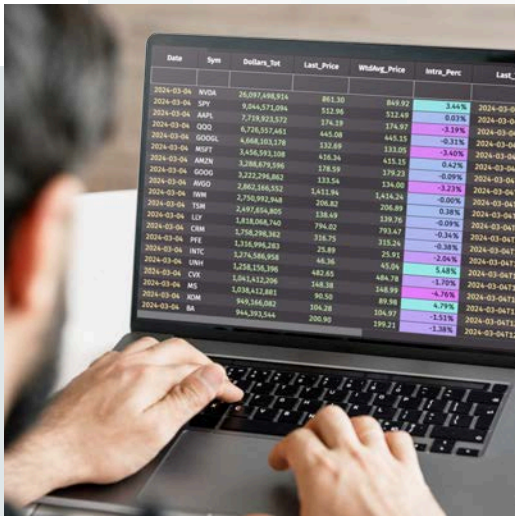
While researching alternative solutions to technologies like Spark/Flink/Dataflow, our engineers spotted Deephaven and immediately found it could be a great solution for Wunderkind. Specifically, the live table architecture built into Deephaven caught our attention.

We dove in and found that it ticked all our requirements. First impressions from performance testing found that running on local workstations,

DEEPHAVEN WAS HANDLING HUNDREDS OF THOUSANDS OF MESSAGES A SECOND WITHOUT BREAKING A SWEAT.

Given that the performance metrics aligned, we dove into the code and found a massive toolbox of functionality that we could directly use, and then ultimately modify to our specific needs as our use cases with Deephaven grew.

The community was the last piece of the puzzle. We have to say that the Deephaven team is one of the best we have ever had the pleasure of working with. Ultimately the solution became clear that we should use Deephaven to handle our aggregations.



Date	Sym	Dollars_Mt	Last_Price	Whaling_Price	Intra_Perc	Last
2024-03-04	NVDA	26,097,498,916	881.30	849.92	3.44%	2024-03-04
2024-03-04	SPY	9,046,571,096	512.96	512.49	6.03%	2024-03-04
2024-03-04	AMZN	7,718,923,572	176.19	174.92	-3.19%	2024-03-04
2024-03-04	GOO	6,776,357,461	445.09	445.15	-0.11%	2024-03-04
2024-03-04	GOOGL	4,668,503,818	132.89	133.05	-3.40%	2024-03-04
2024-03-04	HOFT	3,456,993,108	178.59	431.15	0.42%	2024-03-04
2024-03-04	AMCN	3,388,879,596	486.26	199.29	2.89%	2024-03-04
2024-03-04	GOOD	2,222,296,892	133.54	134.00	-3.21%	2024-03-04
2024-03-04	WGO	2,862,166,552	5,418.96	1,416.26	-0.00%	2024-03-04
2024-03-04	WM	2,750,992,848	208.82	208.89	0.39%	2024-03-04
2024-03-04	LV	2,497,604,805	138.49	139.78	-0.99%	2024-03-04
2024-03-04	CLM	1,818,962,940	794.82	793.47	-0.34%	2024-03-04
2024-03-04	PTC	1,756,286,282	318.75	312.26	-0.38%	2024-03-04
2024-03-04	INTC	1,516,996,283	28.49	28.91	-2.84%	2024-03-04
2024-03-04	PLT	1,276,586,958	48.36	49.06	-1.48%	2024-03-04
2024-03-04	UMH	1,241,156,398	482.85	484.78	-1.92%	2024-03-04
2024-03-04	CHK	1,041,132,358	148.58	148.99	-4.79%	2024-03-04
2024-03-04	HS	1,038,412,881	90.50	89.88	-2.11%	2024-03-04
2024-03-04	KCM	940,346,882	104.28	104.97	-1.38%	2024-03-04
2024-03-04	BA	844,393,544	200.99	199.21	-1.38%	2024-03-04

BEFORE + AFTER ARCHITECTURE

We leverage Redpanda for our event bus, handling 200-250k messages per second at peak, totaling billions of messages a day.

Deephaven handles the entire event load processing and data analysis via Kafka and deployed as statefulsets in a GKE cluster, split out across partitions in the various Kafka topics.

The heap value ranges from 10 to 32GiB, with 1-3 CPUs per pod, each with roughly 25 total instances on average per job.



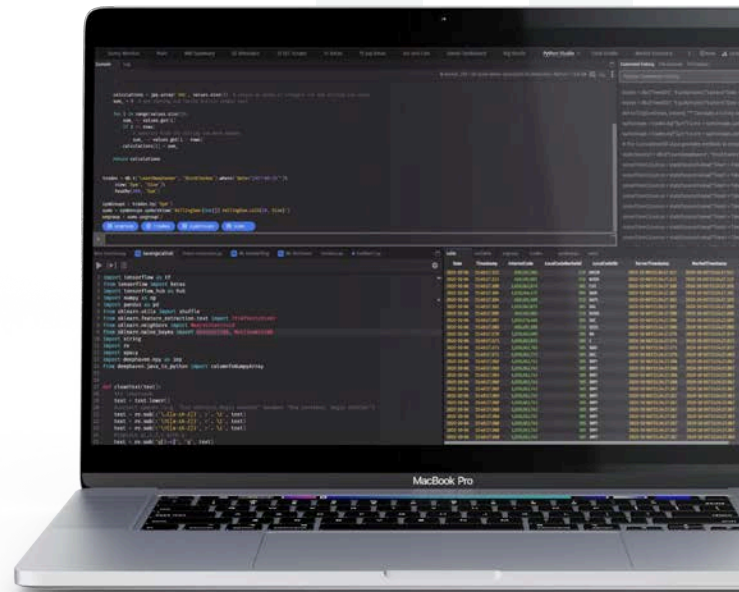
THE RESULTS

We didn't know we needed it, but we now have a full-featured IDE environment for debugging analysis problems and tracing data quality issues. This allows anyone with Python experience to add additional components on the fly.

The new solution also allows for a shift from a bespoke batch-oriented and cumbersome ETL system to incrementally pre-computing analysis in real time as data is collected from real users.

After implementing Deephaven, we saw a

**~20% OVERALL REDUCTION
IN COSTS COMPARED
TO THE SOLUTIONS IN
PLACE PRE-MIGRATION.**



LEARN MORE AT

deephaven.io ↗

