

Captain Metrics Ditches MongoDB for SingleStore: 30,000X Performance Gains Help Get User Journeys Under Control FAST

30,000X
faster than MongoDB

4,000X
faster than PostgreSQL

5M in 1ms
5M-line simple aggregation:
SingleStore 1ms,
PostgreSQL 4sec,
MongoDB 30sec timeout

“We hit the sweet spot with SingleStore. If you cannot do massive analytics, you need SingleStore. If you don't want to manage a clone of your data in a separate database like Google BigQuery, you need SingleStore. If you need to work on fresh data, you need SingleStore.”



Pierre Bazoge
Founder, Captain Metrics

Users are engaged in increasingly complex journeys across multiple platforms/devices/channels including phones, laptops, tablets, and connected TV (CTV). Understanding which of these contribute to a conversion requires effective multi-touch attribution. When companies can accurately understand this, they can determine the right mix to create exceptional cross-device digital experiences.

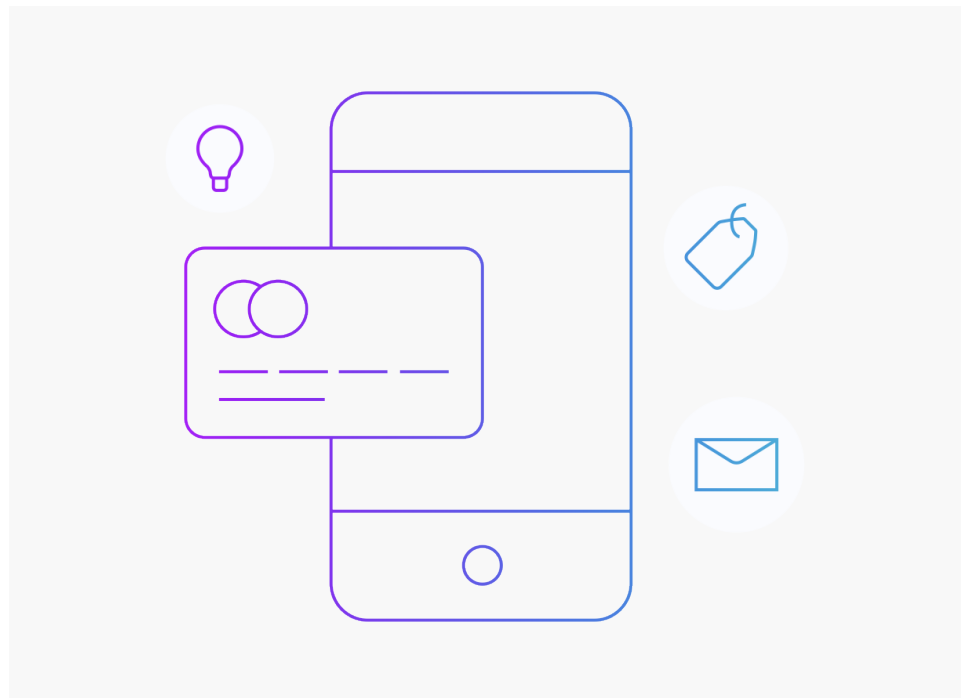


Figure 1: Multi-touch attribution helps companies understand customer journeys

Unfortunately, many advertisers base their digital marketing strategies and investments on false numbers when a sizable share of traffic and interaction turns out to have occurred with bots instead of with actual human users and buyers.

Pierre Bazoge, Founder, Captain Metrics, is a visionary in the industry.

Bazoge discovered the digital marketing reality gap when he reverse-engineered a Google Analytics implementation for a project and found substantial attribution problems. Bazoge decided it was time to build a brand-new platform that would resolve these issues and manage the entire digital marketing cycle. As a result, he founded Captain Metrics as a bootstrapped software-as-a-service (SaaS) disruptor in 2019. Captain Metrics offers a cloud marketing platform that unifies

and enriches customer data to personalize and automate digital marketing across all channels in real time. The platform's main features include a customer data platform, analytics, marketing automation, affiliate marketing, and an AI-driven recommendations engine.

Bazoge explained Captain Metrics' core competitive differentiator this way: "Real-time is rare in the customer data platform (CDP) industry, but that's where the value is because it unlocks many insights and capabilities." And while real-time may be rare among CDPs, real-time has been required in martech/adtech for years. So Bazoge and Captain Metrics are offering CDP users their only real shot at keeping pace with the millisecond-speed industry that is digital marketing.

"Real-time is rare in the customer data platform (CDP) industry, but that's where the value is because it unlocks many insights and capabilities." — *Pierre Bazoge, Founder, Captain Metrics*

"The goal here is to collect and unify all customer data into the central database in order to analyze the marketing efforts and activate these data in real time. Most of the data come from web activities, sessions, users, page views, server-side historical synchronizations, and webhooks, said Bazoge.

Challenges/Goals

While it moves fast to provide always-on user experiences, Captain Metrics also goes deep to provide a comprehensive 360-degree, privacy-respecting view of user customer data and conversion history, including:

- Unification and segmentation of customer data
- Real-time personalization and marketing automation
- User-friendly interface so marketers don't have to ask developers and data scientists for the split-second insights they need
- Customer data deduplication and modeling
- Real-time connection of all data sources via dozens of API integrations
- AI-powered, real-time, one-to-one cross-sell and upsell product recommendations

With Captain Metrics, users can create targeted user micro-segments that match key criteria, such as demographics and firmographics (the latter of which are known as "demographics about organizations"). They can then analyze the performance of each marketing channel by each user segment.

Bazoge explained why real-time access is critical to Captain Metrics: "We can apply many filters on data. Because we have a moving dataset, any time users perform actions, they can enter or leave segments. The dataset is moving so quickly that we cannot pre-allocate the data. That's why we need to compute in real time, because our end users will play with the reports in every way possible to extract value from the data."

Captain Metrics used MongoDB for the initial database supporting the platform. It pulled JSON payloads into a MongoDB cluster, and then did aggregations to answer these queries.

“We believed it when MongoDB said that if you do big data and have scaling problems, and you want to do real-time, you should use MongoDB. I was working with MongoDB for many years, which is why I was ready to deploy it inside Captain Metrics,” said Bazoge.

After a few months of testing with a beta client, Captain Metrics encountered a big problem: the dashboard couldn’t load any more data and queries would time out after 30 seconds. It was costing \$500 per cluster per month for the client — and the database was no longer answering simple queries.

Captain Metrics was told that the dataset being computed in the 5GB query was too big to be crunched by the CPU in anything less than 30 seconds. MongoDB explained there were no solutions to the problem, as the index was covered and the schemas were denormalized to optimize for the queries. “It’s not safe to denormalize too much, but at the same time, if you want performance on MongoDB, you have no choice but to do that,” added Bazoge.

“It’s not safe to denormalize too much, but if you want performance on MongoDB, you have no choice but to do that.” — *Pierre Bazoge, Founder, Captain Metrics*

Sharding did not fix the issues, either. “The problem with sharding is that we’re talking about 5GB of data. You don’t do sharding with gigabytes, you do it with terabytes. Another downside of MongoDB sharding is that many aggregate operators in the aggregation pipeline are not available anymore. We would lose a lot of power from the query engine and still not fix our problem,” said Bazoge.

“The problem with sharding is that we’re talking about 5GB of data. You don’t do sharding with gigabytes, you do it with terabytes. Another downside of MongoDB sharding is that many aggregate operators in the aggregation pipeline are not available anymore. We would lose a lot of power from the query engine and still not fix our problem.” — *Pierre Bazoge, Founder, Captain Metrics*

Technical Requirements

“We had a lot of technical challenges to solve, but the idea at the end is to be able to compute the performance of every touchpoint in our customers’ marketing budget,” explained Bazoge.

Cross-device reconciliation is the process where Captains Metrics reconciles customer behavior that starts on one device and converts on another. The platform detects that this is the same person, and then merges all the related activities together.

“95% of the players in the CDP industry don’t do real-time. Tracking these triggers in real time is very important for marketing automation. To be able to do that, when we ingest data, we sometimes need to update or delete rows. This happens the most when we do cross-device reconciliation” – which goes to the heart of multichannel attribution.

Captain Metrics needed a high-performance database to support its data-intensive application that would deliver:

- Operational real-time UPDATE/DELETE/TRANSACTION functionality
- Fast analytics that leveraged RAM + CPU parallelism
- High scalability through sharding
- High availability with replication
- SQL compatibility for business intelligence (BI) access
- An open source or generous free tier for testing and to support its bootstrapped SaaS disruptor

Why SingleStore

Captain Metrics considered many database technologies on its quest to move on from MongoDB:

MySQL

“MySQL was out of the game right away because it could not do parallelism. One query is one CPU. If your dataset is too big, you’re dead,” said Bazoge.

“MySQL was out of the game right away because it could not do parallelism. One query is one CPU. If your dataset is too big, you’re dead.” — *Pierre Bazoge, Founder, Captain Metrics*

PostgreSQL

“PostgreSQL has the mechanism of workers, and it can sometimes activate itself in order to parallelize some data. However, you don’t know when and you don’t know how. It’s not made for the kinds of queries we’re talking about. For doing real-time analytics the way we do it, it’s not the best player here,” Bazoge explained.

Cassandra

"Cassandra is a great column store and it's a great technology," said Bazoge, "but I was so tired of NoSQL databases that I could not jump into another one right away. MongoDB made me lose two years."

"I was so tired of NoSQL databases that I could not jump into another one right away. MongoDB made me lose two years." — *Pierre Bazoge, Founder, Captain Metrics*

Google BigQuery and ClickHouse

"BigQuery and ClickHouse are append-only databases. You can ingest and load batches into it, but it's unable to update or delete transactions to compute stats in real-time. Both solutions are very costly and not very fast," said Bazoge. Another problem with BigQuery was that when you import a dataset into BigQuery every few hours, the data isn't fresh, he added.

Snowflake

"Snowflake is expensive and would not allow Captain Metrics to easily offer in-house deployments to its customers," said Bazoge, "as they would have to enter into negotiations with Snowflake."

"BigQuery and ClickHouse are append-only databases. Both solutions are very costly and not very fast. Snowflake is expensive and would not allow us to easily offer in-house deployments to our customers, as they would have to enter into negotiations with Snowflake." — *Pierre Bazoge, Founder, Captain Metrics*

VoltDB

"VoltDB is the closest to SingleStore so far, but we could not try it. There is no free tier and the documentation is a bit old," Bazoge explained.

Discovering SingleStore

“When we found out about SingleStore, we knew it was more than enough,” said Bazoge. SingleStore is the world's fastest cloud database for data-intensive applications. Captain Metrics ran database comparison tests with a five-million-line basic request that covered how many distinct users it had, with dozens of aggregates, sums, and median metrics computed on those five million lines. The results were stunning:

- MongoDB timed out in 30 seconds
- PostgreSQL responded in four seconds, but Captain Metrics planned on the dataset growing far beyond five million lines; four seconds today would translate to 60 seconds in two months' time
- SingleStore responded in ONE MILLISECOND

“Was there a hack somewhere? It couldn't work this way!” exclaimed Bazoge. “We decided to play with the time dimensions to query more data and make sure there was no catch. SingleStore still responded in less than one millisecond.”

“Was there a hack somewhere? It couldn't work this way! I decided to play with the time dimensions to query more data and make sure there was no catch. SingleStore still responded in less than one millisecond.” — *Pierre Bazoge, Founder, Captain Metrics*

SingleStore's parallelism was the key to this performance. “SingleStore saturates all the CPUs of your cluster to answer your queries. If you have 16 CPUs in the cluster, it uses all of them in parallel to crunch the data and answer as fast as possible,” he explained. SingleStore also keeps data in memory to ensure no time is wasted fetching this information from the disc.

Another feature Captain Metrics appreciated was SQL compatibility. “SingleStore is like MySQL on steroids,” said Bazoge. “You're at home and you know how it works, but it offers much better performance than MySQL. It's also compatible with Google Data Studio and other third-party data scientist tools. They can connect directly to our database to provide value to it.”

“SingleStore is like MySQL on steroids. You're at home and you know how it works, but it offers much better performance than MySQL. It's also compatible with Google Data Studio and other third-party data scientist tools. They can connect directly to our database to provide value to it.” — *Pierre Bazoge, Founder, Captain Metrics*

SingleStore DB's free tier was also large enough for Captain Metrics' bootstrapped SaaS platform. “This is also important because we're talking about a proprietary database that has a very large free tier. It was large enough to give us time to scale our business,” he explained.

Bazoge also had high praise for SingleStore support. “The team at SingleStore, its documentation, and the forum are very nice and helpful. These resources gave me the confidence to bet on SingleStore. Now it pays off because to me, it's the best technology for doing what I'm doing, and, as a result, we are going in a very good direction.”

Solution

Captain Metrics now has a solid foundation for its data fabric based on SingleStore, Google, and Cube.js.

SingleStoreDB Self-Managed runs on Google Compute Engine

Captain Metrics is using [SingleStoreDB Self-Managed](#) self-hosted on Google Compute Engine. It has one master aggregator and two leaves. Each has eight CPUs, 32GB RAM, and SSD storage. “My customers can basically query one year of data with just these two machines. That’s quite cost-effective: if you want to do the same with other types of databases, you will need many more machines because they don't parallelize cores,” said Bazoge.



Figure 2: Captain Metrics' real-time customer data platform (CDP) architecture

Captain Metrics works by using a data collector for ingestion. A queuer parses the data and controls the rate limits. From there, it pushes the data into SingleStore. An API server transforms the raw data, inserts it into the correct tables, and makes sense of the data.

An analytics server using Cube.js works alongside this workflow. Cube.js allows Captain Metrics to describe schemas and dimensions. It converts the API request into a SQL query and handles the joins without the user writing the queries manually. It creates the SQL, runs the query, and returns properly parsed JSON documents.

The automated processes of Cube.js combined with the ultra-fast query performance of SingleStore has made for a winning combination at Captain Metrics.

Captain Metrics has now Open-Sourced its product to solve for the complexity of the industry's legacy data stack

Captain Metrics already offers users something new and powerful: a real-time CDP. Now it is taking its offering to the next level.

The most prevalent current data stack in digital marketing involves stitching together Airbyte, Snowflake, Google BigQuery, and data science tools to reverse ETL the massive dataflows required to compete in the space. Companies need to hire new people and expand opex further by adding more and more tools. It is costly, time-intensive, and complex, and relies on rigid data schema.

“We are building a new Marketing Operating System for digital marketers.”

— Pierre Bazoge, Founder, Captain Metrics

So Captain Metrics has now Open-Sourced its product to replace all of that cost and complexity with a single unified solution for data collection, modelization, transformation modernization that breaks through that data silos that have separated these functions. “What we are doing is building a new Marketing Operating System for digital marketers,” said Bazoge. “Everyone is doing something unique and we are addressing that in a way that adds flexibility and ease of access to plug whatever apps they want into one unified data fabric. Our solution provides 90% of what they need and leaves 10% free for their own customization, using their own tools, algorithms, and tables. We provide a framework for them to add their own business logic into apps.” The Captain Metrics solution competes with Adobe and other digital marketing products while providing enhanced user segmentation, analytics, and many other capabilities.



Captain Metrics

Outcomes

30,000X faster than MongoDB and 4,000X faster than PostgreSQL

During the database evaluation process, SingleStore answered a five-million-line simple aggregation in one millisecond, while MongoDB timed out entirely in 30 seconds and PostgreSQL took four seconds. For more complex queries, SingleStore continued to shine. A typical 60-day multi-touch attribution report by Captain Metrics involves 53 operations across three tables with six JOINS in three nested levels. It only took 1.8 seconds for SingleStore to respond to this 4.5GB, 5.3-million-line query.

Infinite scalability

SingleStore is built with a scalable architecture, so Captain Metrics can continue to use the same database technology as it grows without running into performance bottlenecks. SingleStore enables the company to scale its data infrastructure, power real-time analytics, and ensure that it has the backend stability needed to deliver its SaaS platform. "SingleStore's scalable architecture guarantees us that we won't need to replace it anytime soon," said Bazoge.

Greatly simplified tech stack and improved productivity

"SingleStore allows us to do highly performant real-time analyses on our data without having to duplicate the data into BigQuery. It saves months of engineering time to have only one database, along with being able to update/delete rows inside the database and query analytics in those same rows," Bazoge explained.

Going from NoSQL to a SQL-compatible database offered additional infrastructure simplification.

"Going back to the SQL world was like a rebirth because I've been working with NoSQL databases for many years. I forgot how great it is to make sure that your entity is saved in only one table and then joins will do the rest. And transactions are able to connect to Metabase, MindsDB, Cube.js, and Tableau software without any problem," Bazoge explained.

"Going back to the SQL world was like a rebirth because I've been working with NoSQL databases for many years. I forgot how great it is to make sure that your entity is saved in only one table and then joins will do the rest. And transactions are able to connect to Metabase, MindsDB, Cube.js, and Tableau software without any problem."



Pierre Bazoge
Founder,
Captain Metrics

Hitting the database sweet spot

Captain Metrics looked at many database options, and SingleStore hit the sweet spot for many reasons, including:

- Its ability to handle operational analytics at any scale
- No need to export data to an exotic database for analytics
- SQL-compatible driver for BI
- The capability to update and delete data in the column store

New AI-powered Use Cases

"Another important thing for Captain Metrics is future new use cases. Because SingleStore is a database that can handle a lot of queries in parallel and complex queries with big datasets, you can run AI inside it," said Bazoge.

Captain Metrics can plug in SingleStore partner [MindsDB](#) to run AI models directly inside its SingleStore cluster. The result: transactions, analytics, AI models, and AI model results all within a single database! This enables the platform to run AI queries on fresh data for anomaly detection, eCommerce recommendations, and similar functionality.

"I'm not paid to be a SingleStore Ambassador. I talk for free and have nothing to gain. I just love the technology and want to help make it bigger."



Pierre Bazoge
Founder,
Captain Metrics

Moving Towards an Open Source Customer Data Platform Future

"The future of Captain Metrics is to open-source it," declared Bazoge. "We believe that every major tool in the data science, data-intensive application industry will get an open source alternative sometime soon that will disrupt everything. In two to three months, Captain Metrics will be an open source customer data platform that does multi-touch attribution and cross-device reconciliation, for free, based on SingleStoreDB."

Bazoge believes in SingleStore's technology so much that he became [a SingleStore Ambassador](#). "I'm not paid to be a SingleStore Ambassador. I talk for free and have nothing to gain. I just love the technology and want to help make it bigger."



Captain Metrics

Watch Pierre Bazoge describe how he and Captain Metrics are using SingleStore

to help companies understand their user journeys and conversions [on this webinar](#), Captain Metrics: Why We Ditched MongoDB



We are honored to feature Pierre Bazoge as an industry-respected SingleStore Ambassador on our [SingleStore Developer page](#)



SingleStore is helping companies compete and win across many verticals. [Learn More >](#)