

Octane

Octane Achieves <200ms Response, Scales to 100s of terabytes, and Lowers TCO by 25% with SingleStore on Google Cloud

10-100M

10s to 100s of millions of measurements per day

100s of TBs

in the dataset

<200ms

response time for real-time analytics

1+ Million

rows of time series data ingested/day

-25%

TCO with SingleStore

“Octane is one of the first real-time billing systems able to ingest real-time data at massive, reliable scale and with high throughput. This capability is exploding in the current market, taking us to places our competitors have never been, and SingleStore helped us achieve that.”



Karan Dhabalia
Founding Engineer, Octane

Octane is a leading usage-based billing platform built for software-as-service (SaaS) product and finance teams to support evolving monetization strategies. The platform provides the flexibility to price and bill based on any metric powered by real-time data, with simple integration to accounting tools and payment providers.



IMAGE / Octane offers usage-based billing for product & finance teams at SaaS companies

Challenges/Goals

Traditional SaaS billing software requires users to aggregate all the data on their side and then figure out pricing at the end of the month. It requires a lot of maintenance and infrastructure to support, and doesn't allow SaaS companies to explore more dynamic monetization strategies.

“SaaS companies don't want to pay flat fees. We want to help them align revenue with their customer's value. As their customer drives more value from the product, the company can charge them more for the added value they are providing,” explained Karan Dhabalia, Founding Engineer, Octane. “Usage-based billing helps companies do that.”

With deep backgrounds in building SaaS products at scale, Octane knows the importance of pricing flexibility, billing simplicity, and getting the right tools to the right stakeholders. When SaaS companies switch to usage-based billing, they can unlock a good boost in revenue and increase their net revenue retention.

CASE STUDY

Octane

“We want to make it easy for product and finance folks to iterate on pricing without having to go back to the engineering team for data. It becomes possible for those two functions to operate independently and at scale,” said Dhabalia.

Octane only requires a one-time setup process for engineers to integrate data into the system. They don’t have to continually check on or maintain it.

The platform has real-time systems that ingest user data and provide continuously updated revenue dashboards. However, queries were very slow, especially on data that spanned across time, resulting in an analytics experience that didn’t meet business user expectations.

“The real-time revenue dashboards becoming slow was a big headache, especially for users who are regularly checking these analytics,” explained Dhabalia.

The database Octane was using at the time, a popular time series database based on PostgreSQL, was not able to scale up to meet user demand. To try to speed up performance, the Octane team had to create many indices on top of the database. However, this approach came with several drawbacks.

Writes became slow because committed writes had to be written to every single index. Data storage also became bloated, because that data was copied in multiple indices.

“We were running huge clusters with a lot of storage for data that wasn’t really that large. We were storing 10X the amount of data we needed because of all the indices we had on top of the database,” said Dhabalia.

“We want to make it easy for product and finance folks to iterate on pricing without having to go back to the engineering team for data. It becomes possible for those two functions to operate independently and at scale.” — Karan Dhabalia, Founding Engineer, Octane

Another challenge related to the database’s slow performance involved Octane’s real-time alert system, which triggers when users cross certain usage thresholds.

“We want the alerts to be as real-time as the data that users are sending us. But if the processing or queries on our side take longer, then we can’t send an alert until it’s done, even if the user sent the data 10 minutes ago,” said Dhabalia.

Potential cross-selling and upselling opportunities were lost due to that lengthy delay. Horizontal scalability was also an issue. “If we wanted to add more nodes, we just had to beef up the machine. But we couldn’t scale horizontally and use a distributed database to help us scale our queries faster,” said Dhabalia.

Technology Requirements

The existing combination of the time series database and Kafka for message-passing to achieve high throughput was not keeping up with Octane's growth. The team started to look at alternative database solutions to better support the platform.

The ideal solution would have:

- Extremely high-throughput reads and writes
- Real-time processing on 10,000 measurements per second
- Online analytical processing (OLAP), online transaction processing (OLTP), hybrid transaction / analytical processing (HTAP) capabilities in a unified database
- SQL support, especially around aggregating queries on time series data
- The ability to meet compliance requirements for sensitive billing data, offering consistent, secure, and transactional reliability

“We weren't really trying to migrate away from PostgreSQL and doing aggregate queries on time series data. We wanted the database to be SQL and transactional because we operate on sensitive revenue data,” said Dhabalia. **“At the same time, we needed it to be extremely performant for analytical workloads because we want to be able to do analytical queries without sacrificing our transactional query loads.”** — Karan Dhabalia, Founding Engineer, Octane

Technology Requirements

Octane considered several other databases in its search for a solution that could keep up with the pace of business.

CockroachDB + Snowflake Added Unwanted Complexity

“We thought maybe we could go with a highly reliable transactional database like CockroachDB and pair it with a data warehouse like Snowflake for analytical queries,” said Dhabalia. “However, we're a lean startup team. We don't want to maintain two data sources. Having to make sure the pipelines that continuously push data from one to the other are in sync seemed like a headache.” Octane did some experimentation on the Snowflake idea with a proof of concept, but taking on low-value work maintaining two data sources was a dealbreaker.

Dhabalia and the team also looked at streaming database options like Materialize, but that solution fell short on predictable, reliable performance because it didn't support ANSI SQL.

Octane

Technology Requirements (CON'T)

Octane Unified its Data Strategy with SingleStoreDB

Then Octane discovered SingleStoreDB, a real-time, unified, distributed SQL database. “In terms of benchmarks, SingleStore seemed like it could do both types of workloads fast. It offered hybrid workloads, transactional processing, analytical data, and high throughput writes combined with horizontal scalability and cost-efficiency,” said Dhabalia. “With SingleStore, we can infinitely scale without having to worry about overloading the system.”

Other SingleStoreDB Cloud features Octane leverages include:

- [Unlimited \(Bottomless\) Storage](#)
- [Pipelines](#)
- [Limitless Point-in-Time Recovery](#)
- [Workspaces](#)

Octane found [SingleStore Support](#) responsive and helpful as the team started using its database technology, including help optimizing how data should be imported and laid out. “We met with SingleStore’s head of product to discuss the potential product roadmap and incorporate our use cases and requirements into the upcoming product roadmap.”

“With SingleStore, we can infinitely scale without having to worry about overloading the system.” — *Karan Dhabalia, Founding Engineer, Octane*

Solution

“SingleStore is our core data store. It powers all our analytics dashboards and ETL pipeline reads and writes. We’re using it for both transactional and analytical workloads throughout our system,” said Dhabalia.

Octane opted for SingleStoreDB Cloud, the fully-managed, on-demand cloud database service, running on Google Cloud, to power data-intensive applications. The team uses two SingleStoreDB clusters; one is a production cluster and the other a development cluster. They are considering expanding to a third cluster.

The primary ingestion pipelines are Kafka-based, used for message passing, and that data is then written to SingleStore.

Python, Go, SQL Alchemy, and GORM make up the rest of Octane’s technology stack and talk directly with the database. The team does not use any off-the-shelf business intelligence tools. Instead, it writes queries on top of SingleStoreDB Cloud.

Octane

“We receive millions and millions of rows a day of time series data. We’re dumping it into SingleStore, running real-time analytics on it, then also doing the transactional, reliable billing process. It’s a hybrid model,” said Dhabalia.

“SingleStore is our core data store. It powers all our analytics dashboards and ETL pipeline reads and writes. We’re using it for both transactional and analytical workloads throughout our system.” — *Karan Dhabalia, Founding Engineer, Octane*

While Octane’s queries do not have many joins, they are complex in different ways. “We don’t have too many joins, but we do have a lot of group-bys. A lot of aggregations in group-bys, and a lot of time series aggregate functions like latest max-min, first-last sort of thing in several group-by clauses,” explained Dhabalia.



IMAGE / Octane lets companies free their pricing strategies from their engineering backlog

Octane

Outcomes

Since choosing SingleStoreDB Cloud to act as its primary unified data store and power its real-time analytics dashboards, Octane is enjoying these and other benefits:

Scalability to Support Future Growth and 100s of TBs of Data at 25% Lower TCO

Octane is no longer constrained by the horizontal scaling limitations it suffered with its original database solution. While the platform's data currently sits in the one-to-10 terabyte (TB) range, the team anticipates stabilizing at 100s of TB of data. While Octane is operating at a higher scale than it did with its previous solution, it's also saving, lowering its total cost of ownership (TCO) by 25%.

"We can now onboard customers that have very high real-time throughput requirements. We definitely have customers sending us 1,000s of requests per second. Before we adopted SingleStore, we had to be careful with onboarding those customers, but now we don't," said Dhabalia.

Achieving High Transactional Writes and Analytical Reads on 10,000+ Measurements per Second

SingleStoreDB Cloud's unified data engine for transactional and analytical workloads powers Octane's fast, real-time analytics platform. Sensitive billing and revenue data are consistent, secure, and transactional for reliability. Meantime, real-time usage data from Octane customers can come into the system at very high throughput with 10,000 measurements per second.

"We're receiving millions and millions of rows a day of time series data. We're dumping it into SingleStore and running real-time analytics on it, and we're also doing our actual transactional reliable billing process on it," said Dhabalia.

Helping Business Users Act Quickly on Their SaaS Revenue Data through Real-time Dashboards with <200ms Response Times

"Realistically, for the analytics dashboard, we're looking for sub two seconds time. A couple of seconds for those is no problem. In our ETL pipeline, it's critical we don't have any lag. Two hundred milliseconds at the most. We have a small set of queries, but each query operates on a lot of data and there are a lot of applications," explained Dhabalia.



IMAGE / Octane's scalable platform ingests 10s to 100s of millions of measurements per day

Continuing to Break New Ground in Real-Time SaaS Billing

"Octane is one of the first real-time billing systems able to ingest real-time data at massive, reliable scale and with high throughput," said Dhabalia. "This capability is exploding in the current market, taking us to places our competitors have never been, and SingleStore helped us achieve that. We're going to support multi-region and multi-cloud in the future, since SingleStore offers support for those capabilities. We also use read replicas for analytical workloads, and SingleStore Workspaces supports that."

Octane plans to help users take a deep dive into price experimentation with back testing, forecasting, price matching to different customers and cohorts, and other ways of helping SaaS companies get the most value out of usage-based billing. "We now have a reliable, fast database that's necessary for us to do modeling and algorithmic analysis on that data with SingleStore," concluded Dhabalia.



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