

BIGCOMMERCE

How BigCommerce Upleveled Kafka Management for Digital Retail Innovation



Headquarters

Austin, TX

Industry

Retail technology

Challenge

Difficulty scaling architecture with open source Apache Kafka® to meet real-time demands of merchants. Resulting in over 20 hours a week managing low-level Kafka infrastructure and not enough time spent innovating new products and features.

Solution

Switching from self-managed Kafka to Confluent Cloud gave BigCommerce a cloud-native, fully managed data streaming platform that is cloud agnostic with pre-built enterprise features.

Results

- Time and cost savings, industry-leading reliability with a 99.99% uptime SLA, and seamless integration with third-party systems—in short, more innovation with far less effort.
- In just five months, BigCommerce migrated 1.6 billion events a day to Confluent, and saved 20 hours a week in Kafka management, along with the scalability savings gained from no longer having to overprovision clusters to accommodate seasonal traffic spikes.

When it comes to smooth e-commerce capability in the modern era, BigCommerce provides the technology *behind* the technology—the platform that's ultimately the backbone of buying for tens of thousands of merchants, including Solo Stove, Skullcandy, Yeti, and a lot of other incredibly popular brands and products.

For a company that enables varying levels of demand for merchants in the digital age, timeliness and accuracy of data are crucial. Inventory has to be accurate to the millisecond, and when data is delayed by a day, or even a minute, it results in lost revenue for merchants and damages to the organization's reputation.

That's only one of the reasons BigCommerce pursued Apache Kafka for their data streaming use case. The modern, open SaaS, e-commerce platform works with tens of thousands of merchants who serve millions of customers around the world. Its digital commerce solutions are available 24/7, all year long, for B2B, B2C, multi-storefront, international, and omnichannel customers.

As Mahendra Kumar, VP of data and software engineering at BigCommerce, puts it, "Real-time data is critical for current business. The value of data drops almost exponentially as time goes by. The sooner you can act on data, the more valuable it is."

Recognizing the need for real-time data while understanding the burden of self-managing Kafka on their own led BigCommerce to choose Confluent—allowing them to tap into data streaming without having to manage and maintain the data infrastructure.

"We started wondering, can we offload all of the management of Kafka—and still get all of the benefits of Kafka? That's when Confluent came into the picture."

— MAHENDRA KUMAR, VP OF DATA AND SOFTWARE ENGINEERING, BIGCOMMERCE

The challenges of self-managed Kafka when instant access to data insights is critical to business

Before Confluent, BigCommerce was managing its own Kafka cluster that was growing and required increasing maintenance. BigCommerce started out with six broker nodes, but as their traffic and use cases increased, they ended up trying to manage 22 broker nodes and five ZooKeeper nodes. This required over 20 hours a week dedicated to maintaining and managing Kafka infrastructure and about half a full-time engineer's time.

They chose Kafka because it is a robust, resilient system designed for high throughput, which would allow the company's data teams to retain data for a period of time, as well as offset and reprocess data as needed. However, open source (OSS) Kafka presented significant limitations. "It was fine initially," says Kumar, "but as the volume of data started getting bigger and bigger, we got to a point where managing Kafka became almost like a full-time job."

Rather than being able to focus on delivering new services that improve user experiences, team members were bogged down by software patches, blind spots in data-related infrastructure, and system updates.

Challenge #1: Maintaining and managing Kafka

"We had to constantly patch software, and from time to time we would see traffic surging and have to constantly manage the scaling of resources. It was a painful journey."

— MAHENDRA KUMAR, VP OF DATA AND SOFTWARE ENGINEERING, BIGCOMMERCE

Having the latest and greatest improvements to Kafka is important, and BigCommerce wanted to get the most out of the infrastructure that was running Kafka. However, managing upgrade priorities had to be balanced with other roadmap features that the team needed to deliver. "If we have to spend time on Kafka upgrades and patching, then it will take time away from our other roadmap priorities," said Kumar.

Once the company got up to 22 Kafka nodes, management became unwieldy. The engineering team was composed of nine people to manage the data platform, data applications, data APIs, analytics, and data infrastructure. However, they had three engineers spending half their time on Kafka-related maintenance and upgrades.

Before big shopping events like CyberWeek, the entire team needed a month to get the technology ready to scale and had to overprovision their clusters by 10-15% to account for the uncertain traffic volume. This exercise was expensive and time-consuming, and valuable resources were being squandered because of the unknowns.

Challenge #2: Inability to tap into real-time data creates burden for customers

"The next evolution of our use case in Kafka: taking each critical type of e-commerce event and building a data model which could load data in real time using a stream processor."

— MAHENDRA KUMAR, VP OF DATA AND SOFTWARE ENGINEERING, BIGCOMMERCE

Kafka was originally instituted to help BigCommerce capture critical retail events such as orders and cart updates. These events were stored in a NoSQL database as they came in. But with that model, the company was not able to tap into streaming data, and that meant they couldn't exactly provide real-time data analysis for merchants who wanted to be able to extract quick insights from the data.

From time to time, the system would go down, and that would cause a spike of events to backlog, which degraded the merchant and end-user experience. BigCommerce had an ETL, batch-based system for analytics and insights, resulting in a wait time of eight hours before merchants could get their analytics reports. This system consisted of 30+ MapReduce jobs. On occasion, jobs would fail and need manual intervention, causing even further delays.

"Overall, the merchant experience has definitely improved. It's been seamless, and the spikes and delays are no longer there."

— MAHENDRA KUMAR, VP OF DATA AND SOFTWARE ENGINEERING, BIGCOMMERCE

The decision to adopt a fully managed Kafka solution

At the time of migration, BigCommerce processed around 1.6B events a day. These were high-traffic, high-value events such as order events, cart events, and page view events. The team knew that with Kafka, they already had the ability to consume data in real time. They just needed the support of a fully managed data streaming platform to help them tap into the potential of Kafka. As Kumar puts it, "The people who built Kafka are the people at Confluent, so we were already familiar with the company."

In addition, the company's storefront architecture was already on Google Cloud, so the fact that Confluent Cloud integrates so easily with GCP was a key lever. Using Confluent on Google Cloud allowed BigCommerce to significantly cut down on data transfer costs.

But Kumar's team wanted the platform they chose to ultimately be cloud agnostic, in case they ever decided to change vendors or add another cloud. Confluent met that criteria, too. "The combination of Confluent on Google Cloud was the perfect choice," says Kumar. "It provided the best of both worlds."

Last but not least, BigCommerce needed the platform to enable them to stream data to their own customers—the merchants—as part of their valuable open SaaS strategy. Kumar explains, "We wanted to enable our merchants to get their commerce data into their own data warehouses so they could create reporting and combine it with their other kinds of data." This would enable merchants to integrate with third-party systems such as Google Analytics and other APIs to, for instance, run ad campaigns for custom audiences.

Migrating from self-managed Kafka to Confluent Cloud

When BigCommerce wanted to migrate from their existing open source Kafka to Confluent Cloud, they embarked upon a phased approach to ensure a smooth transition without any impact to their critical workloads.

Kumar's team meticulously planned the migration with the help of Confluent's expertise to design the partition strategy and migration solution. This was done to prepare for future needs and traffic variations for production use cases.

The first phase of the migration saw the piped production data being sent to both OSS Kafka and Confluent to monitor latency and avoid any possible downtime. Once the systems were fully in sync, the switch to Confluent was seamless.

In the second phase, BigCommerce integrated consumer lag and sink connector lag reporting into a single unified Google Cloud monitoring dashboard with the help of Confluent Metrics API. This helped them keep track of the lagging issues and provided a way to address them in a timely manner.

The third phase involved the migration of insights and data lake workloads, loading data into Google BigQuery using Confluent Sink Connector. Finally, in the fourth phase, BigCommerce optimized Kafka stream performance and determined the optimal number of consumers for future workloads.

In five months, BigCommerce had migrated 1.6 billion events per day, 22 broker nodes, three clusters, and 15 TB of storage data to Confluent Cloud. They experienced zero downtime, zero data loss, and zero disruption to the merchant experience.

"Confluent is designed for scale and resilience. We get all of these benefits, too, which makes it even more appealing."

— MAHENDRA KUMAR, VP OF DATA AND SOFTWARE ENGINEERING, BIGCOMMERCE

With Confluent Cloud and fully managed Kafka, attention turns to innovation

The resulting benefits of migration for BigCommerce's engineering team have had a strong impact on the business, the merchant, and the end-user experience.

The biggest benefit is the simplification of using Kafka. With Kafka fully managed in [Confluent Cloud](#), the engineering team no longer has to spend 20 hours a week managing low-level Kafka infrastructure—upgrades, traffic spikes, compliance, security, nodes, and other tasks. They also haven't had to overprovision clusters by 10-15% "just in case," because Confluent is elastically scalable. Instead, they have been able to turn their full attention to innovating products in a highly competitive industry.

"That's been huge on the business side," says Kumar, "because it frees up engineers from worrying about operations so they can focus on functionality, like, for instance, our new Meta integrations," which allow merchants to use BigCommerce Kafka streams to feed data right into Meta's Conversion API and create ad campaigns to highly customized audiences in real time.

In addition, Confluent enables the team to design products and features around asynchronous communication—critical because, as Kumar says, "The world is moving toward a more disconnected one where every system can publish and consume messages. Some systems consume immediately, and some consume at other times. The ability to publish and consume messages asynchronously will be central to any good systems we build."

For merchants, the switch to Confluent means they can get real-time analytics on BigCommerce's open platform and tap into precise insights for their specific use cases. The data coming out of BigCommerce is now trustworthy, timely, accurate, and actionable. Because Confluent is cloud native and has the ability to "connect everywhere" via integrations, BigCommerce is able to take real advantage of the data pipeline.

They're now pulling in data via the flexibility and ease of [Confluent's Google Cloud Storage Sink Connector](#), and merchants can easily combine this real-time data with the data they already have in their ecosystems so they can get to the ideal state of a 360-degree view.

The challenges of building a [real-time solution](#) on Kafka are not always easy to solve, particularly for an in-house engineering team that's better off spending time on product innovation versus tech-stack management. As Kumar explains, "If you're a company looking at building a real-time solution and managing your own open source Kafka, there is going to be a lot of overhead. Some costs are visible; some are hidden." Kafka self-management requires people, time, and effort—and can be a process of trial and error.

"In our case," says Kumar, "Confluent makes so much sense. We no longer worry about Kafka clusters or scaling. We just focus on product development."

"The system we decided to go with is a real-time messaging system (Kafka) backed by a managed provider (Confluent) on a reliable, robust, performance cloud platform (Google Cloud)."

— MAHENDRA KUMAR, VP OF DATA AND SOFTWARE ENGINEERING, BIGCOMMERCE

Business Results:

Reduced operational burden

The team no longer worries about managing Kafka and instead focuses on feature functionality. Specifically, there's no longer a need for the data team to spend 20 hours a week managing low-level Kafka infrastructure.

Faster time to market

As artificial intelligence (AI) and other technologies become competitive levers for retail companies, BigCommerce can now get things like a product-recommendation proof of concept built faster.

The ability to design for asynchronicity

Confluent enables BigCommerce to build products and features that enable every system to consume messages in real time or not.

Cost savings

The simple fact that BigCommerce no longer has to assign engineers to manage Kafka creates cost savings, specifically around the operational overhead that had previously been required to scale, patch, upgrade, monitor, rebalance, and optimize the platform.

Technical Results:

Elastic scalability

Confluent has the ability for high throughput, so it handles traffic seamlessly, even during seasonal superspikes like CyberWeek. There's no need to overprovision clusters by 10-15% "just in case" like they used to.

No downtime

Since completing the migration to Confluent Cloud, there has been zero data loss and zero downtime.

Integration with third-party systems

The ability to connect streaming data from Kafka with [Confluent's many connectors](#) provides great engineering flexibility and opportunity.

Learn More About BigCommerce

Read BigCommerce's [Product and Engineering blog](#)

Follow BigCommerce [Product and Engineering on Twitter](#)