

STORYBLOCKS.COM

# Storyblocks Partners with Confluent to Redesign Its Data Pipeline and Slash Technical Debt



## Storyblocks

### Headquarters

Arlington, VA

### Industry

Media

### Challenge

Create an event bus to improve data collection, event-driven communication between microservices, and data discovery for new use cases.

### Solution

Confluent Cloud creates a seamless flow of data across systems and empower developers to take full advantage of streaming event-based data.

### Results

- Reduced technical debt and TCO from REST APIs and modernized a monolithic application
- Consolidated search for audio and visual files
- Support of rebranding effort
- Vastly improved data visibility, collection, and communication for data scientists and developers

Ranked the fourth-fastest growing media company in the U.S. by Inc. Magazine, Storyblocks was born out of a desire to solve the problem of a lack of accessible, affordable stock for creatives. It's the first unlimited download subscription-based provider of stock video and audio, claiming over 100,000 customers in the television and video production industry, including NBC and MTV, plus tens of thousands of hobbyists looking to enhance their video projects and productions. Storyblocks' subscribers can download an unlimited number of clips from a vast and rapidly growing library of stock video, production music, motion backgrounds, sound effects, special effects, and more.

As the company transitioned from scrappy disruptor to major industry player, it began to experience issues with a monolithic application they had built when the company was started. They also experienced other issues later with synchronous REST API calls between services, after they had split the monolith into microservices. Namely, developers and data engineers couldn't resolve issues fast enough or iterate on their search functionality with sufficient agility, and were taking on a significant amount of technical debt as an increasing amount of data threatened to slow down the productivity and time to market for new features.



## Technical Solution

For the first version of their re-architected stack, Storyblocks initially used an AWS Kinesis data pipeline dumping raw data into Amazon S3, which enabled certain machine learning features and analysis of clickstream data. At this point, they'd already, in parallel, built a central pipeline of events and had begun to decouple their microservices and were moving on from focusing on user accounts to setting up an event-driven microservice for billing. But when they began trying to use this newly established pipeline for inter-service communications, it started to break down and fail for a variety of reasons, including a lack of scalability, and not really being a suitable solution for fully decoupling services.

To manage these challenges, the Storyblocks engineering team needed a new solution that could form the backbone of an entirely new data pipeline—one that could enable change both internally and externally and provide a foundation from which to digitally transform Storyblocks' various teams and services by giving them a level of data speed, agility, visibility, and inter-communicability they'd never had before.

They began to use Apache Kafka® and did a proof of concept in conjunction with using a schema registry from Aiven. After about six months, though, they realized they weren't getting the support for Kafka that they needed from Aiven, so they started to use Confluent Cloud, a fully managed cloud service for Kafka. Simultaneously, they began trying to spread the use of Kafka as an event bus for more streaming applications and machine learning (ML) features, and they realized Confluent Cloud very quickly made a difference.

"For us, just having one way of communicating was game-changing," says Charles DeVeas, Director of Engineering, Storyblocks. "Having the one event pipeline that's flexible for everyone's use case that we can use to communicate or to absorb data, that's amazing. In a way, Confluent is its own microservice, the microservice of data collection and data communication maintained by its own team. If we didn't have that, people would be building REST API endpoints left and right and it would cause massive entropy."

"Confluent is definitely the central pillar of our data infrastructure," continues DeVeas. "All data collection that isn't an application database is going on our central pipeline through Confluent. We have about 200 topics and they all have important transactional data, from billing to clickstream data and more. We're getting engineers and data scientists to use it as a communication method and to leverage historical data for new use cases."

With this Confluent-backed data pipeline in place, the Storyblocks team could begin to affect a true digital transformation both internally and externally.

Instead of implementing a queue for inter-services communication, the team just puts it on the pipeline where it's stored forever. This sort of infinite storage is powerful for two reasons:

1. Events can be replayed on demand with powerful in-built schema validation.
2. According to DeVeas, "The infinite retention is important because analysts sometimes look at historical data when answering various questions for the business; I'll bet that at some point in the next six months some squad or team in the business has new questions about our content contributors' historical behavior, and an analyst will ask to dig into our content service events."

## Business Results

**Reduction of TCO and technical debt.** "It can be really hard to quantify technical debt, but as an anecdote, we had this huge problem with a bug in our billing service after we implemented it using Kafka, and it did affect users because it cancelled them early or changed their end date incorrectly, and it affected about 5,000 users," says DeVeas. "The normal way this would play out is customer service would find out about it, go on Slack, and say 'hey all these people are complaining.' Engineers would go to the production database, write SQL queries to figure out the problem, and then have to write a migration, maybe do some cowboy queries in the production database to fix the information. And maybe it gets resolved but you end up with this disjointed solution where the engineers had messed with the production database and there was a separate fix implemented in the code, and the whole thing would take maybe a week depending on the bug. But with Confluent, we were able to fix the bug with just a unit

test. And because the bug was implemented as part of an event consumer we didn't even have to go to the database and have engineers mess around with the database. All they had to do was fix the bug with a unit test, maybe over the course of two days, and then just replay the events and by replaying the events we confidently knew everything was fixed and validated the results."

**Rebranding support.** "One of the most noteworthy benefits of using Confluent for us has been for our rebranding project, which was the end result of our microservices project to consolidate all of our services into one Storyblocks.com where users could search videos and audio in the same place. Users used to have to log in to two different sites and now they can just log in to one site and get access to the audio and video in one place. Confluent was a big part of this."

**Improved search.** "We're a search product for artists and other types of people who need certain types of footage quickly, so obviously our search algorithm is fundamental to our business. We wouldn't be able to iterate on our search algorithm as quickly as we can now if we didn't have Confluent as that central data bus. We're now able to easily collect all of the search data where people hover over and download the videos, so the user search experience in 2022 will be greatly improved as we begin to use more machine learning."

## What's Next?

For the future, DeVeas envisions broader use of an events-based architecture at Storyblocks to improve the user experience and to tackle the lucrative enterprise market.

"Engineers will be building more features using events as opposed to using custom API calls from service to service," says DeVeas. "And we want to launch more ML-based features, meaning more batch features on the backend for video analysis and more user-facing real-time features, and both are going to use Confluent. I definitely expect to see a quadrupling of machine learning features and hopefully a doubling of engineers using it for communication in 2022. For example, we want to use machine learning to convert more enterprise users by finding out if their behavior is consistent with an enterprise customer."

## Learn More About Storyblocks

<https://storyblocks.com>

*"Confluent is definitely the central pillar of our data infrastructure. All data collection that isn't an application database is going on our central pipeline through Confluent."*

— Charles DeVeas, Director of Engineering, Storyblocks