

UC San Diego Reduces Time-to-Value for Core Business Processes with Integration Layer Built on Confluent

UC San Diego

Challenge

Reduce friction in core business processes to enable rapid real-time responses to student, staff, IT admin, and researcher demands

Solution

Use Confluent and Apache Kafka® as the centerpiece of a university-wide integration platform

Results

- Time-to-value improve
- Streaming workloads rapidly scaled
- Storage costs lowered
- Development time and costs cut by 80%

In light of the global pandemic, The University of California, San Diego had to shut down on-campus facilities, resulting in a loss of millions of dollars every month. The university felt immense pressure to transition and build out their online learning platform, ultimately transforming how they serve their students, employees, researchers, and IT admins. However, their existing legacy architecture did not allow the IT team to be agile enough to anticipate and react quickly to real-time changes. Addressing these challenges meant rethinking how IT looked at real-time data. Fortunately, the university had recently launched the Enterprise Systems Renewal (ESR) program, its largest technology modernization initiative ever. This program is aimed at streamlining core business processes and transforming how the university leverages real-time data.

A primary goal of ESR was to reduce friction within and among business processes—from budgeting and bidding on research to onboarding students and hiring employees—that relied on an aging IT infrastructure. For UC San Diego, the key to eliminating this friction was a new integration-platform-as-a-service (iPaaS) middleware layer based on Confluent, the platform to set data in motion. This became the first step moving toward a new event-driven architecture that would serve their legacy modernization efforts over the next 10 to 15 years.

This event-driven architecture was pressed into service almost immediately in the school's coronavirus response. This included the deployment of a testing app to integrate test results with student data, which would guide decision-making for when students returned. With the new iPaaS event streaming infrastructure in place, UC San Diego built and deployed this testing app 12x faster. "With the middleware layer we had built using Confluent, we tackled a challenge that could have had a substantial impact on the university's bottom line. We built and deployed a testing app in two weeks, streaming test result data in near real time," says Scott Lee, Enterprise Architect at UC San Diego. "Within a month we had a working, production-ready application that would have taken a year to build out in the past. Our middleware layer—with Confluent at its core—improves time-to-value, and that's critical to our operational model and invaluable to the business."

UC San Diego is already seeing its event-driven architecture investment pay dividends, with an even greater return on investment expected as the campus community becomes increasingly mobile. "Today many workflows are based on the desktop—signing into Oracle Financials or PeopleSoft from a workstation, for example. In the future,

Within a month we had a working, production-ready application that would have taken a year to build out in the past. Our middleware layer—with Confluent at its core—improves time-to-value, and that's critical to our operational model and invaluable to the business.

— Scott Lee, UC San Diego

with 5G, we see that more and more workflows will be mobile, and that's where we'll rely even more heavily on our streaming architecture," says Lee. "The event streaming framework we have with Confluent enables us to build for the future. When we see a threat on the horizon we can mitigate it and when we see an opportunity we can take advantage of it." UC San Diego is now processing

90 million records a day from their Learning Management System, opening up new revenue streams for other UC campuses—including the UC Office of President, and tracing student health via a wastewater heatmap to detect early traces of the COVID-19 virus across the campus.

Business Results

Time-to-value improved. "In the era of IT, time-to-value is a key differentiator for any organization," says Lee. "We invested in a middleware layer that has event streaming with Confluent at its core, and now we're reaping the benefits of it in terms of time-to-value. That's why we were able to deliver a mission-critical testing app in weeks, instead of months or a full year."

Streaming workloads rapidly scaled. "In just nine months, we went from no workloads on our middleware layer to 100 and we went from no streaming workflows in production to more than 150," says Lee. "Still, in many ways we are just scratching the surface of our usage of Confluent and Kafka."

Storage costs lowered. "Right now, our Splunk spend is significant, as is our budget for in-memory storage with SAP HANA," says Lee. "With ksqlDB we are able to bring that under control by filtering data in-stream and delivering more curated data to our data warehouse."

Development time and costs cut by 80%. "Our iPaaS model has opened new opportunities for us, as many other universities are interested in adopting it," says Lee. "In fact, by sharing our iPaaS and setting up some new Kafka topics, we enabled UC Merced to reuse our student activity hub. We did this in about a month, and it costs about \$20K per month to run, whereas we estimate building it out from scratch would have taken six months to a year, and would require \$100K per month to operate."

Technical Solution

Prior to undertaking the ESR program, UC San Diego business processes were often bogged down by mainframe and other legacy systems that hampered the ability to recalibrate business processes as customer needs shifted or new ones emerged. "Much of our data was locked in legacy data warehouses, and accessing it required navigating a gauntlet of processes," says Lee. "As a result, management lacked the telemetry to make sight-line adjustments. They were swivel-chairing between applications that were not integrated, and they had very little access to the real-time data they needed."

When Lee and his team began putting together the iPaaS infrastructure to address these challenges, open source solutions—including Kafka, Apache NiFi, and Apache Airflow—played a central role. So too did considerations for a strategic move to the cloud. "We want to be out of the data center business and we want to go full cloud. Many of our vendors are already making the move," Lee explains. "For example, Oracle has Oracle Cloud Financials and Quali is offering their higher ed and research software on AWS. However, we also want to avoid cloud vendor lock-in that solutions like AWS Kinesis can lead to."

The choice of Kafka for event streaming enabled the team to pursue a build-it-yourself approach in which they built and owned their infrastructure and data operations. They built activity hubs for finance, research, students, and employees. By designing these hubs to blend data via their new iPaaS, it became possible to combine research data with finance data or with student or employee data in real time when and as needed.

As developers got started with Kafka, their initial efforts were not always well-aligned with best practices. This led the group to engage consultants from Slower for support. "Because they were using legacy methods, our developers weren't doing things in the most efficient way. We wanted to build our own pipelines, but we saw that to stand up the core capability we would need an expert sourcing model. We had talent, but we needed true experts," recalls Lee. "That's where Slower—and later Confluent—come into the picture. They helped us get off the ground and get to where we are now, with other universities asking us how they can replicate our model."

In building out iPaaS applications, UC San Diego developers used numerous Confluent-supported source and sink connectors for Kafka to pull data in from and push data out to existing systems, respectively. "We have the ability to connect to virtually any source and get the data out and put it into Kafka topics," says Lee. "That's very powerful for us."

Over time, more personally identifiable information, financial records, and other sensitive information began flowing through Kafka, underscoring the need for security and monitoring. "As we bring in more private data, the security and monitoring capabilities of Confluent Platform are becoming more and more important," says Lee. "We are very bullish on using Confluent Control Center to improve the monitoring of our pipelines, from source to target."

Having already demonstrated the time-to-value improvements made possible by its iPaaS layer, the team is now working on optimizing data flow patterns and retiring difficult-to-maintain change-data-capture processes.

Use of the platform continues to grow as interest in replicating UC San Diego's model spreads among other universities. "Last spring, we had relatively little demand on our learning management systems. Then the pandemic hit and everything had to move online. Now we have 90 million records per day flowing from those systems through Confluent workloads," says Lee. "If you look at surveys of university administrators, you'll see that schools throughout the country are looking for real-time student, employee, and learning data. At UC San Diego we already have it. Other schools want to reuse our model, and that's a business opportunity for us."

As we bring in more private data, the security and monitoring capabilities of Confluent Platform are becoming more and more important. We are very bullish on using Confluent Control Center to improve the monitoring of our pipelines, from source to target.

– Scott Lee, UC San Diego

Learn More About Enterprise Systems Renewal at UC San Diego

esr.ucsd.edu