



E-BOOK

# Conquer Your Data Mess With Universal Data Products

A Technical Executive's Guide For  
Data Products Mastery with Data Streaming

# Contents

Click the page number on any page if you'd like to return to the Table of Contents.

## 1 The Origins of the Data Mess

## 2 The Business Impact of the Spaghetti Architecture

## 3 Applying a Data Products Mindset

## 4 Break Down Data Barriers With a Shift-Left Approach

## 5 Universal Data Products Require Data Streaming

## 6 A New Platform for the Universal Data Products Paradigm

## 7 Unlock New Use Cases With Universal Data Products

## 8 Drive Business Value With Confluent's Data Streaming Platform

## 9 Join Leaders Across Industries

# Executive Summary

LEADERS ACROSS ORGANIZATIONS, INDUSTRIES, AND COUNTRIES intuitively believe that data is a valuable asset. However, leveraging its value has never been harder – because the data architecture is a mess. Sprawling point-to-point connections within and across the operational and analytical data estates create an insurmountable web of complexity. Teams waste resources developing, deploying, and operating disjointed solutions that each address a piece of your data problems – but never completely solve your data mess.

And, your data mess directly impacts your bottom line:

- Time to market slows down as teams struggle to find the right data sets
- Operational costs balloon as teams waste time holding the web of complexity together
- Risk increases as each new point-to-point connection is another weak link

To pivot to new operating models, launch new products, or enter new markets – **you must conquer your data mess.**

This eBook shows how you can join over 5000 organizations leveraging Confluent's Data Streaming Platform to deploy universal data products that eliminate data access hurdles, transform legacy data architectures, and minimize risk. With Confluent, you can accelerate time-to-market and reduce TCO, while improving ROI by 5-10x.



"At Allianz, we are modernizing our core insurance system by replacing legacy technologies and migrating over 75% of our applications into the cloud. A data streaming platform like Confluent allows Allianz to quickly react, respond, and adapt to ever-changing data in real time, which is key to driving exceptional customer experience, competitive differentiation, and revenue growth."

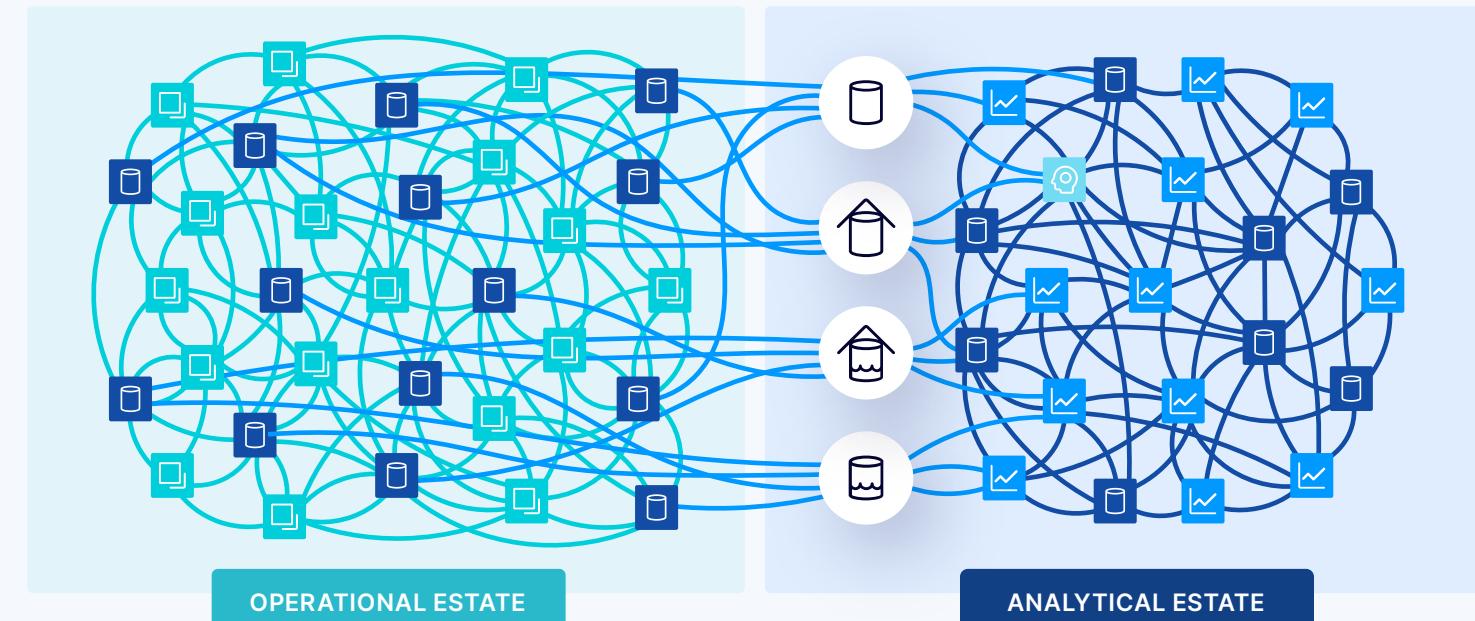
Bruno Fontoura Costa, Head of Global Integration Services and Architecture, Allianz Technology

# 1 The Origins of the Data Mess

Data resides in two different locations across two siloed estates:

- **The Operational Estate – operational databases** that fuel real-time business applications
- **The Analytical Estate – data warehouses, data lakes and lake houses** for after-the-fact analysis to influence business decisions.

In an effort to link data within these estates and transfer data between them, an incredibly complex array of point-to-point integrations and technologies have emerged. But – perhaps, unsurprisingly – these ad hoc solutions have created more problems than they've solved in the long run.



## How did we get here?

Over a decade ago, the data architecture and landscape was rather simple. Operational databases supported transactional systems like payments, processing, inventory, and billing that ran the business. Data warehouses were used to analyze and report on what had already happened in the business, like how much of an item was sold in the last quarter or how much more of that item needed to be restocked.

Messaging systems were implemented for ultra-low latency needs, while ETL / ELT tools moved

data from operational stores to analytical stores in period batches. In some cases, entire business units invested in costly, custom engineering projects, stitching together disparate systems with glue code.

This data architecture was built to meet the relatively straightforward needs of businesses at the time, but it began to take on a life of its own as those needs grew and evolved.

In the modern enterprise, the average data stack has exploded and modular application patterns have emerged with the adoption of best-of-breed tools. Business units now rely on SaaS applications for daily operations. Developers build cloud-based apps with microservice architectures linked to many databases. Analysts and AI/ML developers each use their preferred data warehouses, lakes, or lake houses.

Adding to this complexity, we're no longer just automating human processes—today, software automates other software. Data flows aren't confined to a one-way street from operational systems to analytics anymore. Now, there's intense pressure to loop insights back into operational systems, forcing teams to create a tangled web of new point-to-point integrations with every additional data source and destination.

Organizations keep stitching together a patchwork of legacy tools and processes—batch data transfers, batch-based ETL/ELT systems, off-the-shelf messaging solutions, REST APIs, and custom-built integrations. This approach leads to a sprawling mess of rigid, complex, and expensive point-to-point connections, ultimately shackling teams and preventing them from fully leveraging their data to propel the business forward.

Today, enterprises have more applications than ever and an **overwhelming** amount of data that is siloed and deeply interdependent, but never fully integrated. Simply put: **It's a mess.**

10x

"There are two key challenges with legacy technology. First, the **cost and risk of making changes** to large monolithic applications is prohibitive. Second, you end up with your **data locked up in a particular technology and vendor**, which can be very challenging to evolve."

Stuart Coleman, VP of Engineering, 10x Banking

## 2 The Business Impact of the Spaghetti Architecture

**THIS TANGLED WEB OF SPRAWLING**, rigid, and complex point-to-point connections forms a costly, fragile, and inefficient data foundation that drags down operational performance, stifles innovation, delays data-driven insights, and undermines customer experience, ultimately placing your business at a competitive disadvantage. Over time, this growing complexity doesn't just erode your organization's ability to innovate, scale, and meet evolving market demands, it also leads to escalating operational and hidden expenses, often costing millions of dollars.

On average, developers spend a significant amount of their time on data-centric tasks – designing and maintaining pipelines, integrating

diverse data feeds, ensuring data integrity, and maintaining robust security and compliance measures. This eats into the time they could be spending on advancing the needs of the business. This happens because organizations have treated data as a mere byproduct of their business operations instead of a vital asset, integrating it in an ad hoc manner when they should be enabling its free movement, everywhere.

To overcome this, teams need to start treating their **data as a valuable product** – an asset that's actively managed, curated, and used with purpose. By making this shift, you can harness data strategically to drive innovation, improve customer and employee experiences, and push your business forward.

### Sainsbury's

"Our customer-facing initiatives would stall due to lack of access to data. A lot of our data lived in very old legacy systems which were so critical for the running of the business that no one wanted to touch them because if they stop, the business stops."

Pedro Baeta, Senior Engineering Manager,  
Sainsbury's

### ebay

"When business requirements are complex, the code needed to implement them also becomes complex. With our legacy platform, everything was interdependent, so updating the code in one area often led to unexpected problems in another."

Hudson Lee, Leader of Platform Engineering, ebay Korea

## Common Customer Challenges

01

### Higher Total Cost of Ownership

**Operational inefficiency:** Managing and maintaining numerous custom integrations is costly and time-consuming. The ongoing cost of monitoring, updating, and troubleshooting each connection can add up to millions of dollars in operational expenses.

**Slow time to value:** IT and development teams often spend too much time maintaining these fragile connections, pulling resources away from innovation and strategic initiatives.

02

### Reduced Agility and Innovation

**Longer development cycles:** The complexity of point-to-point connections slows down the introduction of new features or systems. Each change in the ecosystem requires careful adjustments, making it harder to scale quickly and address evolving business and market demands.

**Inflexible architecture:** The rigid nature of these connections limits the ability to experiment with new technologies or adapt to changing business needs, stifling innovation and competitive advantage.

03

### Increased Risk of Failure

**Single point of failure:** Direct dependencies create a domino effect—if one system fails, it can disrupt the entire communication flow.

**Poor fault tolerance:** Ensuring data integrity and message delivery requires complex logic and robust retry mechanisms, and managing schema changes and maintaining backward compatibility increases the risk of data loss or faulty transmission.

04

### Outdated Insights and Poor Customer Experience

**Fragmented data:** Point-to-point integrations often lead to fragmented and siloed data, making it difficult to achieve a single source of truth. This fragmentation hampers data-driven decision-making and can result in conflicting insights.

**Data inconsistencies:** Underdeveloped governance capabilities, combined with the manual and disparate nature of these connections, often leads to inconsistent data quality and insight inaccuracies, eroding trust in the data and limiting its reusability.

05

### Security and Compliance Challenges

**Increased vulnerability:** Each point-to-point connection is a potential vulnerability that needs to be secured, increasing the effort required to keep data safe and compliant.

**Regulatory risks:** Ensuring compliance across a complex, fragmented architecture is challenging, raising the risk of violating regulations like GDPR, HIPAA, FINRA, CCPA, and others, which could lead to legal penalties and damage to reputation.

# 3 Applying a Data Products Mindset

**Data products** are live, refined, fully governed, and ready-to-use data assets that are discoverable, contextualized, trustworthy, and reusable by data consumers for any use case.

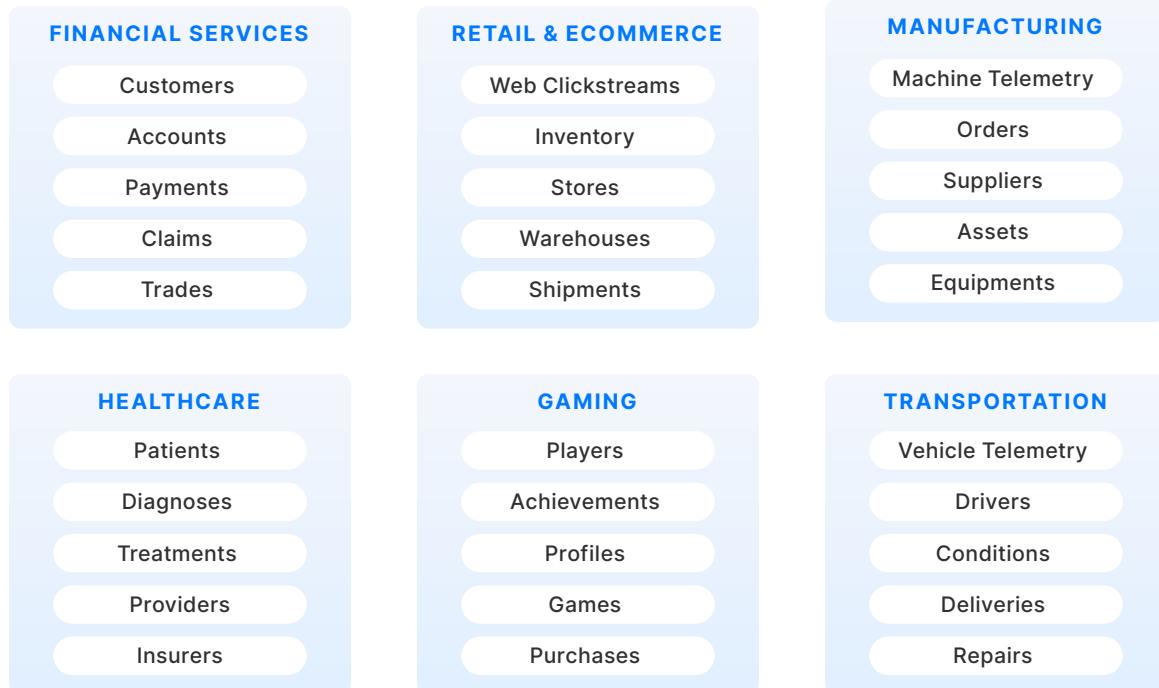
**THE CONCEPT OF BUILDING DATA PRODUCTS** has been around for more than a decade. DJ Patil, the former Chief Data Scientist of the United States Office of Science and Technology, sparked this conversation in his 2012 work *Data Jujitsu: The Art of Turning Data into Product*. Yet, for many organizations, this vision remains unfulfilled. The root of the problem lies in a fundamentally flawed approach – most focus on creating data products solely in the analytical estate.

The problem with this method is that analytical data products use data from the operational estate to function. But the applications, systems, and teams producing the operational data often

don't know how their data is used in analytical data products. When something changes or breaks in the operational estate, the analytical data products that rely on those systems end up becoming untrustworthy, unreliable, and stale. In addition, these analytical data products often fail to deliver value to the operational side of the business, leaving the critical divide between the operational and analytical estates unbridged.

To unlock the full potential of data products across your organization, they need to be universal – discoverable and accessible by data consumers in both the operational and analytical estates.

## Universal Data Products



## From Business Byproducts to Universal Data Products

Imagine each of the important entities in your business – customers, accounts, claims, inventory, shipments, and more. Each of these entities is a data product. Instead of querying data from static tables in a database or data warehouses, each of these data products are live. They are being continuously enriched, governed, and continuously shared so your teams can build applications with trustworthy, well curated data faster, unlocking data value the moment it's created.

When your data is converted into a universal data product, any data consumer can mix and match these data products to solve real business problems. This sparks a **virtuous cycle of innovation**, where each new data product boosts the value of the others and enables more reuse across the organization. With a universal data products mindset, the focus shifts from "Where is my data, and is it accurate?" to "What is my data, and how do I get value from it right now?"

## Deploying Universal Data Products in Your Organization

Harvard Business Review estimates that "companies that treat data like a product can reduce the time it takes to implement it in new use cases by as much as 90%, decrease their total ownership (technology, development, and maintenance) costs by up to 30%, and reduce their risk and data governance burden."

To realize the vision of universal data products, teams must shift away from the current model where they spend excessive time searching, processing, and building bespoke data sets and pipelines. Universal data products must follow these key principles:

- Commit to Data Stewardship:** Teams take ownership of the data they create, contextualizing it for downstream use, fostering collaboration and reuse.
- Embrace Dynamic Data:** Data products undergo continuous updates, allowing teams, systems and applications to act and react to the most up-to-date data.
- Prioritize Governance:** Robust governance should be embedded into the data products, ensuring quality, consistency, and policy adherence, making them reliable and trustworthy organization-wide.
- Bridge the Operational and Analytical Divide:** Operational data must enrich analytics for real-time business insights, while these insights seamlessly influence operational applications.
- Democratize Data Use:** Finally, these data products must be easily discoverable and instantly usable by anyone with appropriate access.

**dish wireless**

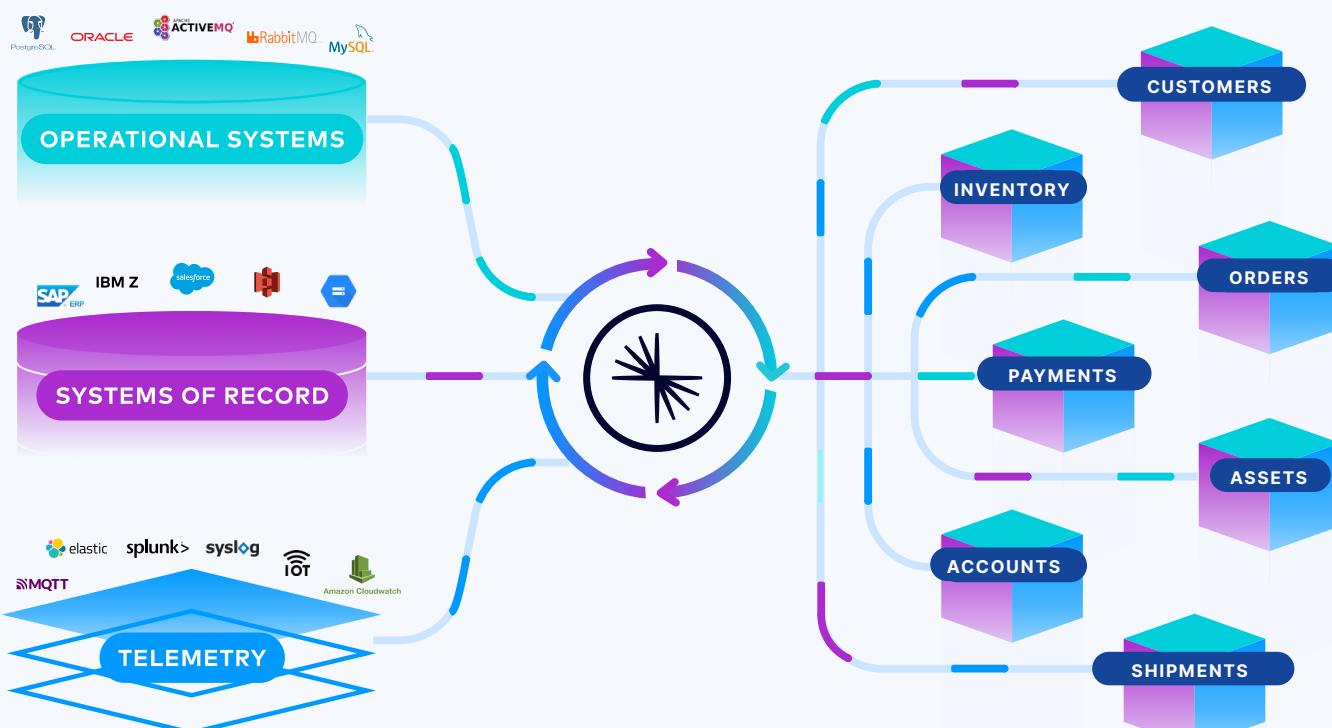
"What I'm most excited about with our partnership with Confluent is our work on data products because data from our network is more valuable than the network itself."

Brian Mengwasser, VP and Head of Marketplace and App Design, Dish Network

# 4 Break Down Data Barriers With a Shift-Left Approach

**IT'S CLEAR THAT** A universal data products mindset is the key to closing the gap between the operational and analytical estates and unifying data across your organization. But what does this mean in practice? The answer lies in changing your approach to where and how you do your data processing and governance – in short, you need to **shift it left**.

The crux of shifting left is relatively simple: moving data extraction, transformation, remodeling, and governance closer to the source, where data is freshest. By combining this approach with a decoupled architecture, downstream consumers gain access to well-defined, consistent, and high-quality data products that are trustworthy and instantly applicable across both operational and analytical estates.



Shift From	Shift To	What It Means
Tightly Coupled Integrations	Decoupled Architecture	Eliminate point to point connections and allow your data producers and consumers to work independently and your data-dependent systems and applications can get the exact data they need, when they need it.
Point-In-Time Query and Processing	Real-Time Processing	Transform data as it's created, branching out new streams to power apps, microservices, and data pipelines in real time. Fuel data products across estates and your enterprise with curated data streams.
Batch-Based Data Integrations	Event-Driven Continuous Data Flows	Let domain owners take charge of data, capturing and sharing events in real time. Store data forever if needed, and replay or reprocess it on demand, ensuring failure resilience.
Downstream Data Wrangling	Governed and Reusable Data	Stop the endless cycle of wrangling and cleansing data downstream. Instead, shift data quality and curation processes upstream. Now, your domain experts can deliver ready-to-consume data that's governed, trustworthy, and reusable, enabling consumers to instantly apply it to their use cases.
Static Centralized Data Estates	Decentralized Data Flows, Shared Everywhere	Get data from wherever it lives to wherever you need it to go — whether your use cases are on-prem, in the cloud, or at the edge.

Embracing a shift-left approach empowers you to build well-curated, governed, and universally accessible data products that effortlessly support both real-time and batch use cases across your organization. However, traditional methods—such as adding more tools to your hodge-podge tech stack or relying solely on centralized data warehouses and data lakes—have fallen short in resolving these challenges. So, how can you effectively address your data mess?

# 5 Universal Data Products Require Data Streaming

OVER THE LAST DECADE, two ubiquitous technologies have transformed the way data is produced, processed, and consumed.

- **Data Streaming:** The constant flow of real-time event data from different sources, so applications can act and react to the data the moment it's created.
- **Stream Processing:** The use of streaming systems/architecture to shape streams of data on-the-fly, enabling in-the-moment contextualization to build streaming applications and unlimited data reusability.

## Data Streaming Is Trusted by Over 75% of the Fortune 500

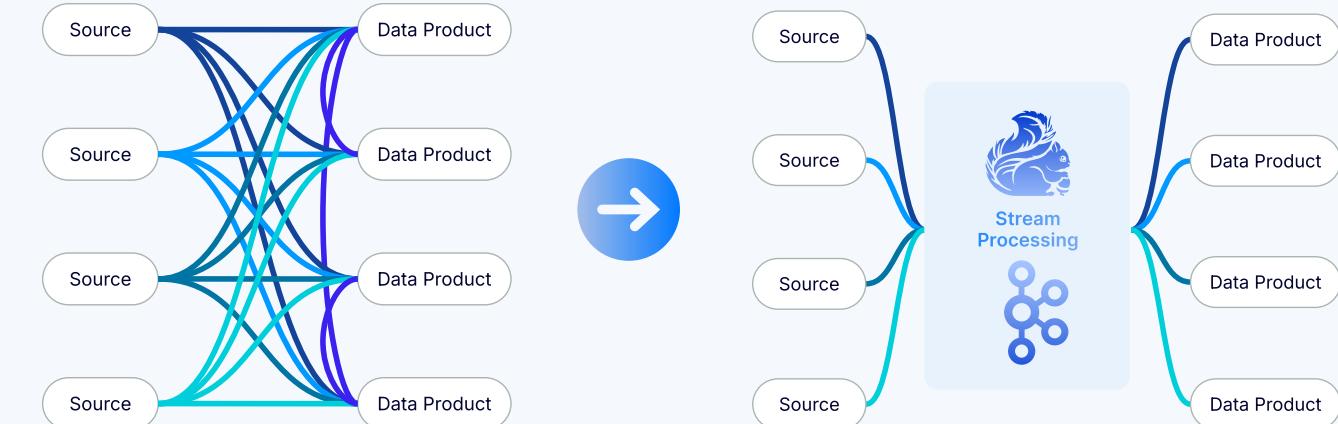
Ten years ago Confluent CEO, Jay Kreps, co-created Apache Kafka® to reinvent the flow of data within organizations through a completely decoupled, highly scalable, fault-tolerant, persistent and event-centric architecture. Kafka lets enterprises reimagine data as something in motion instead of something at rest, supporting the continuous flow of data throughout the estates. Today, Apache Kafka has become the de facto standard for event streaming.

## Stream Processing Is Shaping the Future of Real-time Data Handling

In 2016, a few years after Apache Kafka gained viral adoption across the globe, Apache Flink came into being and today, it has emerged as the open source stream processing standard.

Apache Flink embraces the event-centric paradigm and enables continuous transformation of data the moment it's created, allowing teams to create and share the same data in multiple contexts to downstream systems and applications. The end result is improved data portability, data consistency, and cost savings.

A [scalable](#) and [decoupled](#) architecture as a single [source of record](#) for high-quality, self-service access, to real-time and reusable data products



## The Foundation for Universal Data Products

Together, Kafka and Flink provide the foundation for organizations to create universal data products. However, to successfully implement a universal data products strategy and build a central nervous system for your organization, your teams need a comprehensive **data streaming platform** that goes beyond decoupling point-to-point connections and continuous processing. It must remove the operational burden of managing open source solutions like Kafka and Flink while tackling the critical challenges of data governance, consistency, and quality – all while ensuring self-service access across your organization.

Some organizations try to build a platform that achieves all this, however the chaos of their data mess makes the task insurmountable — wouldn't you rather focus on business outcomes. To implement this platform, your teams must first reimagine their data architecture strategy.

### essent

"Relying on batch processing can cause performance issues and result in poor decision-making based on outdated data. By using Kafka and Flink together in a unified platform, our teams will be able to easily build intelligent streaming data pipelines that can extract data from various sources, process it in real time, and feed it to our downstream consumers for timely analysis without any operational challenges."

Name, Title, Essent



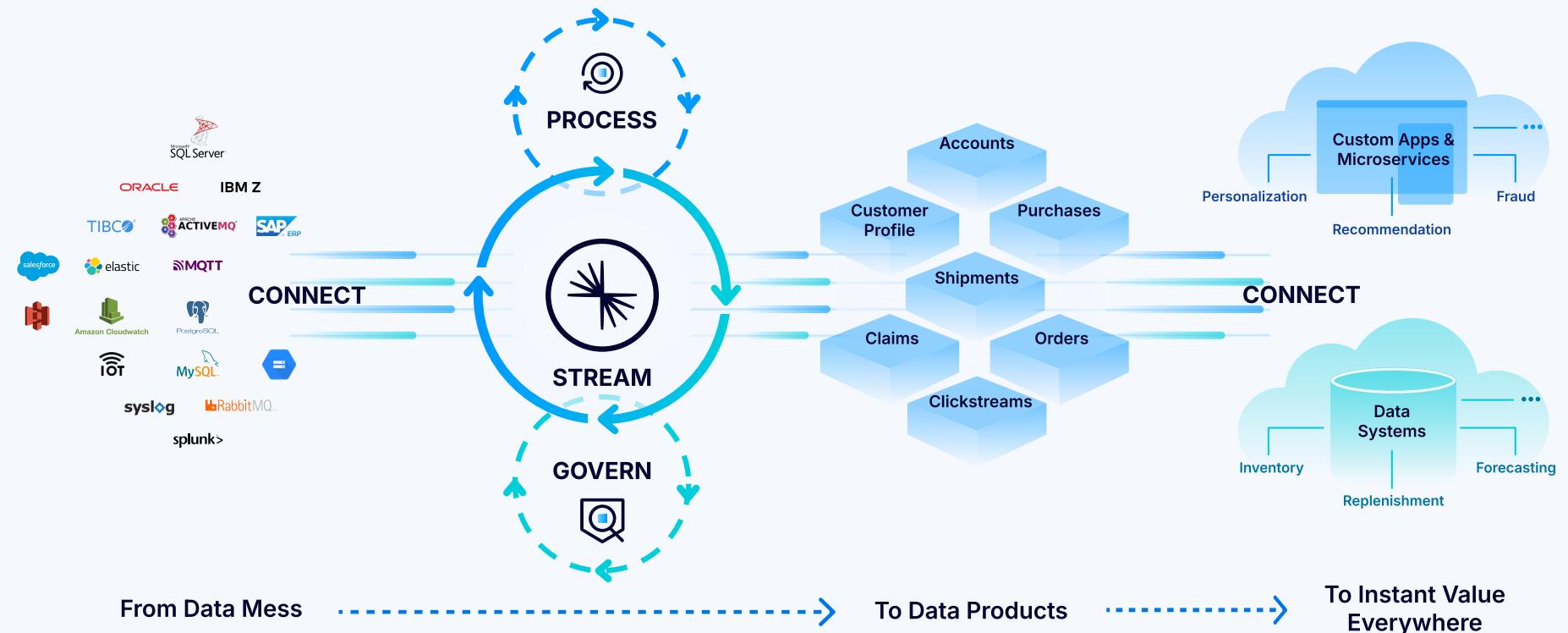
"We wanted to minimize our effort in terms of maintaining Kafka infrastructure so we could focus on building and shipping data products—that's why we chose Confluent. Without a fully managed solution like Confluent, we would have had to dedicate a lot more human resources to running Kafka on-premises by ourselves."

Vinod Chelladurai, Principal Data Engineer, SumUp

# 6 A New Platform for the Universal Data Products Paradigm

THE CONFLUENT DATA STREAMING PLATFORM reimagines your data architecture by building upon the heritage of Apache Kafka and Apache Flink to enable enterprise-wide adoption of data products.

With Confluent, your data architecture is no longer a complicated, expensive, and risky mess. The platform untangles your data problems, breaking down data barriers and silos across your enterprise. It delivers universal data products that connect teams, systems, and applications, ensuring a consistent view of the most up-to-date data.



01

## Stream

Confluent reimagines data streaming through the award winning Kora engine, which removes the need to manage Apache Kafka. With Kora, you have the scale, elasticity, resiliency, availability, and security required for mission-critical workloads across hybrid and multi-cloud environments. Kora improves ROI by reducing TCO through pay-as-you-go storage.

02

## Connect

Data streaming requires connections to be built, deployed, and managed for each data source. Confluent provides 120+ pre-built connectors that allow your teams to instantly connect data. Your teams are free from writing generic integration code and managing connectors.

03

## Process

Confluent lets you join, enrich, and curate data at the source while removing the operational complexity and burden of running Apache Flink. Regardless of whether the data comes from the operational or analytical estate, it's presented in a consistent format that seamlessly works across both.

04

## Govern

Confluent provides a suite of governance capabilities that enable your teams to maintain data contracts, classify and organize data into a neat catalog, track data lineage and securely find and consume trustworthy data products through a self-service portal, while ensuring observability, compliance and confidentiality of data that's on the move.

### Raiffeisen Bank

By completely decoupling our architecture, enriching data in-flight and standardizing our data into a common taxonomy, Confluent has enabled our teams to think of data as a product that is governed, reused and shared. Confluent is the single mediation point for our modern data flow strategy, enabling us to maximize the usability of data and greatly improving developer agility and time to market."

Thomas Joham, Senior Enterprise IT Architect, Raiffeisen Bank International, Raiffeisen Bank

With these four pillars, Confluent's Data Streaming Platform transforms your data mess into reusable, high quality data products. These data products remove the need for brittle point-to-point integrations, and can be leveraged by every team, in any organization, for operational, analytical, and never-before possible use cases.

Confluent's [2024 Data Streaming Report](#) revealed that 79% of IT leaders cite data streaming platforms (DSPs) as pivotal to realizing business agility. And 93% cite DSPs as the key to overcoming pervasive obstacles like data silos.

# 7 Unlock New Use Cases With Universal Data Products

LET'S CONSIDER A COMMON use case enabled by Confluent's Data Streaming Platform: **real-time payments**.

Through data products, your teams can **connect** systems of record data, mainframe data, payments data, ledger data and **stream** to transaction records, historical payments data, and more.

This data can be **processed** to analyze transaction patterns and create a "customer profile" data product, which could then be shaped to show a current "threat vector" for risk scoring purposes. The same data can be further enriched for in-stream fraud detection and combined with other data points (such as geolocation and account login for context) to trigger accurate real-time alerts for anomalous transactions.

Your analytics team can also **reuse the data** to build a "watch list" data product to train machine learning/AI tools to recognize and predict fraud patterns.



"Today's customers demand rich, personalized experiences, and business operations must be optimized to stay ahead of the competition. We use Confluent as an essential piece of our data infrastructure to unlock data and stream it in real time, with use cases like customer 360, e-commerce, microservices, and more."

Yves Caseau, Group Chief Digital and Information Officer, Michelin

At the same time, your business analytics team may want to fuse this "customer profile" data product with "clickstream" and "customer loyalty" data products to fuel instant, hyper-personalized recommendations.

Through **governance capabilities**, you can ensure each of these high-value data assets are trustworthy as they are being produced and reused.

Each iteration of this process creates a new, fully governed data product that increases the value of the others with no duplication or inconsistency of data sets – and no delay for data consumers. Your teams can curate data products on demand and share them wherever they're needed, with self-service access to scale across the enterprise.



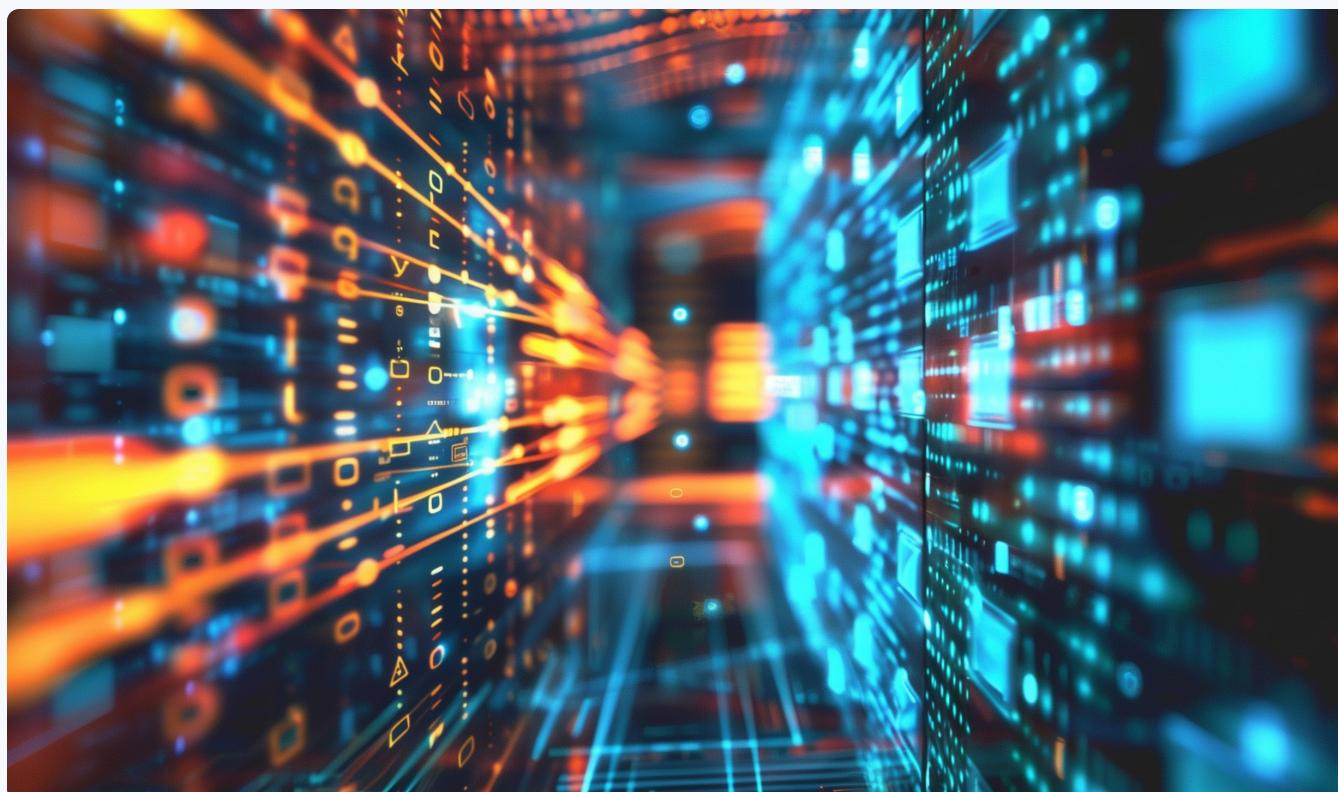
Examples of high-impact use cases that leverage data products include:

- **Fraud Detection and Prevention:** Connect data products for customers, accounts, web logins, and transactions to create a 360° view of every touchpoint in a customer account that can be used to create dynamic threat scores and prevent fraudulent activity.
- **Delivery and Fleet Management:** Bring together data products for vehicle telemetry, drivers, customers, and repairs to streamline how fleets deliver essential services – from route optimization, to preemptive maintenance checks, and everything in between.
- **Customer Loyalty:** Mix and match data products such as customers, inventory, purchases, and web clickstreams, to get a holistic view of customer purchase patterns and build a loyalty program that rewards repeat business.

For more exciting use cases enabled by data products, visit the [use case hub](#).

# 8 Drive Business Value With Confluent's Data Streaming

WITH CONFLUENT'S DATA STREAMING PLATFORM, you're able to transform the way data creates value for your business with new use cases powered by universal data products. You can realize increased team agility and speed, optimize costs, improve your customers' experience, and reduce the risk of security and compliance challenges while increasing scalability. This future-proof data architecture accelerates innovation as you scale to meet evolving market demands while slashing operational inefficiencies that can cost millions.



kakao**games**

"It is crystal clear that Confluent helped us save time to market thanks to the real-time game log analysis. A lot of game publishers today are facing challenges with real-time analysis because the data size is massive and logs often follow a standard. We learned our lessons from the past and Confluent has helped us innovate faster and ultimately enrich the gaming experience."

Eugene Lee, Director of Infrastructure Division, Kakao Games

## Common Customer Benefits

01

### Reduced Total Cost of Ownership

**Resource Optimization:** Remove costly and time-consuming work by eliminating the need to constantly monitor, update, and troubleshoot each point-to-point connection, and save your organization millions of dollars.

**Faster Time to Value:** When there's no need to manage complex and brittle point-to-point integrations, your teams can redirect focus to strategic initiatives. This cuts maintenance costs and prevents technical debt from piling up.

02

### Increased Agility and Innovation

**Rapid Innovation:** Enable real-time, curated, and reusable data access across the organization, speeding time to market for new products and services. Your teams can focus on meeting customer needs, staying ahead of competitors, and scaling with market demands.

**Better Collaboration and Trust:** Foster collaboration across departments with consistent and trustworthy data. Self-service discovery and secure access ensure everyone is on the same page.

03

### Scalability and Future-Proofing

**Adaptable Architecture:** Seamlessly integrate new data sources and destinations with a decoupled architecture and remove point-to-point connections as a single point of failure. Unleash a continuous flow of data between operational and analytical systems to create a virtuous cycle of data-driven innovation.

**Continuous Improvement:** Adapt to changing business needs, ensuring your organization's data architecture is up-to-date, reliable, and resilient over time.

04

### Real-time Insights and Improved Customer Experiences

**Precise Personalization:** Use accurate, real-time data to instantly tailor experiences for your customers, delivering targeted offers, custom recommendations, and relevant content that boosts satisfaction and loyalty.

**Proactive Engagement:** Analyze interactions as they happen to identify potential issues and opportunities. Build stronger relationships with customers by anticipating their needs, resolving concerns, and responding to behavior changes in real time.

05

### Robust Security, Governance and Compliance

**Secured Data:** Maintain centralized standards for observability, security and access controls to ensure data confidentiality and protect data from unsanctioned access, reducing the security and vulnerability risks created by disjointed point-to-point connections.

**Regulatory Compliance:** Avoid legal and reputational damage in the ever-evolving landscape of data regulations by removing one-off solutions and adopting a unified approach to ensuring all data sources and teams meet regulatory requirements.

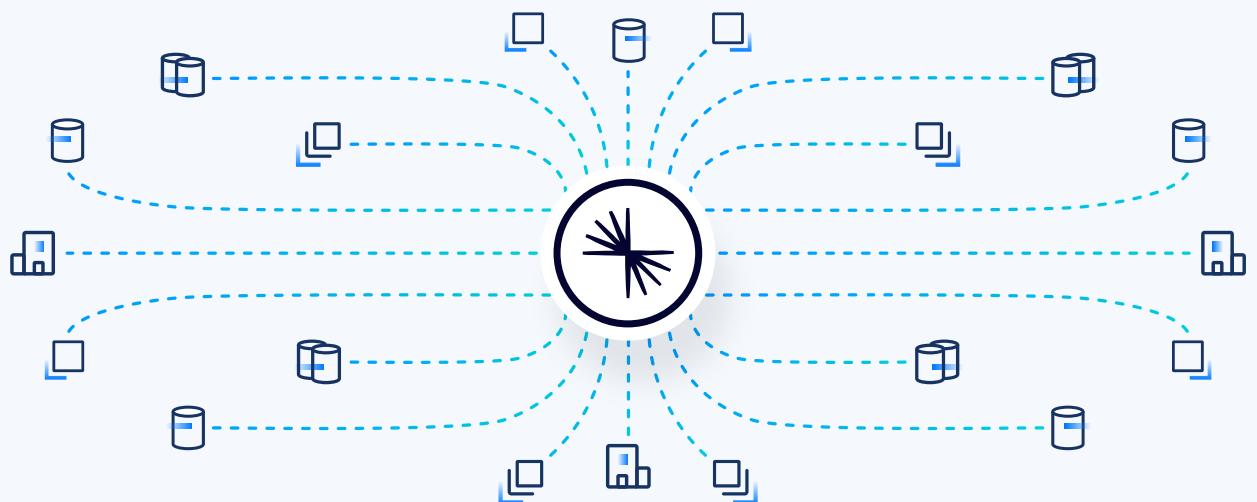
## 9

# Join Leaders Across Industries



## Conquer Your Data Mess

Conquering your data mess is not just a technical necessity—it's a business imperative. With Confluent's Data Streaming Platform, you can transform disjointed data into universal data products, enabling your teams to unlock new value across operational and analytical estates. This paradigm shift not only streamlines your data architecture but also accelerates innovation, reduces TCO, and minimizes risks. Many of the world's most recognized brands are already using Confluent to fuel their teams with real-time, trustworthy data.



FORRESTER®

Learn why Confluent was named a Leader in [The Forrester Wave™: Streaming Data Platforms, Q4 2023](#), with the highest strategy and vision scores across all participating vendors. In the report, Forrester calls Confluent a "streaming force to be reckoned with."



If you had a 360° view of every data event in your company at any time...

- » What problems would you solve?
- » What opportunities would you capitalize on?

Conquering your data mess is not just a technical necessity—it's a business imperative. With Confluent's Data Streaming Platform, you can transform disjointed data into universal data products, enabling your teams to unlock new value across operational and analytical estates. This paradigm shift not only streamlines your data architecture but also accelerates innovation, reduces TCO, and minimizes risks.

See How 4,000+ Tech Leaders  
Are Driving Value With Data  
Streaming Platforms

[Download the 2024 Data Streaming Report](#)



### RESOURCES

- [Turn Data Mess to Data Products with a Data Streaming Platform](#)
- [The Forrester Wave™: Streaming Data Platforms, Q4 2023](#)
- [Capital One's Data Management Tips for the Modern Enterprise](#)
- [Data Streaming Use Cases Library](#)