DeltaStream

The Case for a Stream-First Architecture

Operationalizing Real-Time Data for AI,
Analytics, and Beyond

Cost Efficiency

DeltaStream

Power real-time decisions
and AI innovation with streaming-first infrastructure.

Table of Contents

Introduction

Meeting modern user expectations requires real-time, streaming-first, intelligent operations.

Customers demand immediacy in every aspect of their digital lives—how they shop, bank, hail rides, and engage with online services. These expectations have been shaped by digital-first disruptors building on modern, real-time data platforms. At the same time, generative AI and intelligent automation are resetting the bar entirely, as users expect hyper-personalized, responsive, and context-aware experiences in every interaction.

To meet these expectations, businesses must build intelligent, data-driven operations that respond in real time. That requires abandoning batch-based systems—built for overnight jobs and static reports—in favor of streaming-first architectures that process and act on data the moment it's generated. While many teams have adopted tools like Kafka for microservices, the analytics layer remains stuck in batch. In today's world—where AI agents, fraud detection systems, and pricing engines must operate with millisecond precision, a streaming-first architecture isn't a luxury. It's essential.

A streaming-first architecture treats event streams as the default input for both applications and analytics—enabling data to be processed and acted on as it arrives, without waiting for scheduled jobs or batch refreshes. It reduces operational and compute costs by minimizing redundant batch processing, accelerates time-to-insight through real-time dashboards and decisioning systems, and opens the door to new revenue opportunities via proactive, data-driven applications. From personalized AI assistants that deliver hyper-relevant responses in real time, to fraud detection systems that act within milliseconds, to dynamic pricing engines that adapt to shifting signals, and IoT platforms that process time-sensitive telemetry at scale—stream processing powers the intelligent applications modern businesses rely on.

What Is a Streaming-First Architecture?

A streaming-first architecture redefines the modern data stack by treating every piece of data as a continuous event, not a static file. Unlike batch-first systems, where data must be collected and stored before processing, a streaming-first system processes and enriches data in-flight. Traditional pipelines rely on pre-defined schedules, complex orchestration layers, and redundant data transformations. This approach not only adds operational overhead, but also creates significant lag—making it impossible to respond to changes in real time.

In a streaming-first setup, raw events (from Kafka, CDC streams, or pub/sub systems) are ingested and immediately processed through declarative transformations. Data is filtered, joined, enriched, and routed using familiar languages like SQL, and the resulting streams are simultaneously delivered to downstream consumers like data warehouses or real-time dashboards. This architecture reduces the need for batch orchestration, eliminates the need for costly full refreshes in data warehouses, and powers real-time applications and analytics with fresh, context-rich data at the moment it's generated.

Why Streaming Is Still Hard

Despite the clear benefits of streaming-first architectures, many organizations struggle to adopt them due to technical and operational challenges.

Operational complexity remains one of the biggest blockers. Managing stateful stream processors like Flink or Spark Structured Streaming is inherently more complicated than managing batch jobs. These systems require fine-grained tuning, provisioning compute resources dynamically, setting up and maintaining checkpoints, and handling scenarios like out-of-order data. Ensuring high availability and fault tolerance across distributed infrastructure adds additional layers of difficulty. Unlike stateless batch jobs, streaming systems must be always-on and resilient to failures in real time, which requires sophisticated monitoring and alerting. Without deep operational expertise and tooling, infrastructure management can quickly become overwhelming.

Second, data governance gaps in stream processing are far more pronounced than in traditional batch pipelines. In many Kafka-based setups, there's little visibility into who owns which stream, what schemas are being used, how data is being processed, or how transformations are applied along the way. Schema drift, missing metadata, inconsistent naming, and ad hoc logic make it difficult to understand how data flows from source to consumption. This obscurity creates risk—not just for compliance and auditability—but for operational teams trying to debug issues or build confidently on top of streaming data.
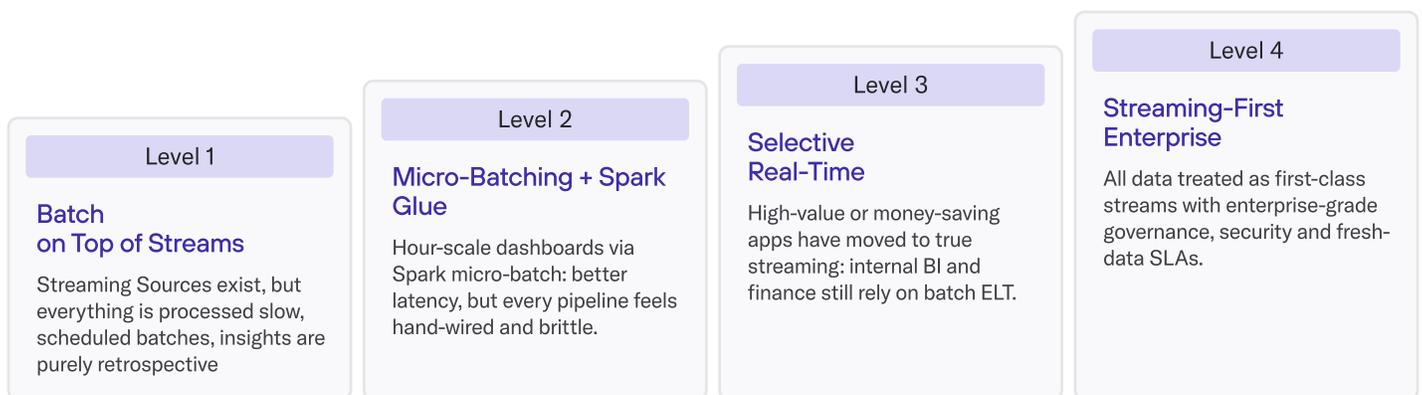
Third, developer bottlenecks often slow progress. In many organizations, stream processing logic is embedded in backend applications, written in low-level code, and deployed alongside critical services. This makes it difficult for data teams to experiment, debug, or modify pipelines without help from engineers. This limits the agility of analytics teams and slows time-to-insight.

Overcoming these hurdles requires a platform that simplifies streaming while enforcing governance and offering self-service tooling to data practitioners.

Streaming Maturity Model

Organizations don't become streaming-first overnight. Most progress through a series of stages—starting with batch-on-streams and evolving toward fully governed, real-time data platforms that support both operational and analytical use cases.

Where do you fall in the streaming maturity journey?

Level 1

Batch on Top of Streams

Streaming Sources exist, but everything is processed slow, scheduled batches, insights are purely retrospective

Level 2

Micro-Batching + Spark Glue

Hour-scale dashboards via Spark micro-batch: better latency, but every pipeline feels hand-wired and brittle.

Level 3

Selective Real-Time

High-value or money-saving apps have moved to true streaming: internal BI and finance still rely on batch ELT.

Level 4

Streaming-First Enterprise

All data treated as first-class streams with enterprise-grade governance, security and fresh-data SLAs.

Level 1: Batch on Top of Streams

In this early stage, organizations have streaming sources like Kafka or CDC-enabled databases in place, but they're still relying on batch pipelines to do the heavy lifting. Instead of processing data as it arrives, teams schedule jobs to extract and land data into warehouses or lakes at fixed intervals—often hours or even days later.

A common example is generating weekly fraud reports by bulk-loading Kafka data into Snowflake or Databricks for processing. While the underlying data is real-time, the architecture strips it of its immediacy. There's no capability to act on data in motion—insights are strictly retrospective, and any in-flight decisions or operational triggers are impossible. The streaming infrastructure exists, but analytics and business logic remain stuck in a batch-first mindset.

Level 2: Micro-Batching + Spark Glue

In this stage, organizations begin to move beyond pure batch by using frameworks like Spark to enable micro-batch processing. This allows data to be processed in smaller, more frequent chunks—typically every few minutes—resulting in dashboards that update hourly or near-real-time instead of daily.

A common example is marketing teams generating hourly campaign performance dashboards by micro-batching Kafka event streams into Parquet files for loading into data warehouses. While latency is improved, each pipeline is hand-wired, brittle, and difficult to evolve. Pipelines in this stage often lack standardization and reusability, making every new use case feel like a ground-up effort. Real-time insights remain confined to narrow, predefined windows. While better than traditional batch, teams remain mostly reactive, and the architecture doesn't yet support scalable, event-driven analytics.

Level 3: Selective Real-Time

In this transitional phase, organizations begin to unlock the full potential of streaming—albeit selectively. High-value, business-critical use cases such as real-time fraud detection, recommendation engines, or log analytics are moved onto native streaming platforms like Flink or Kafka Streams. These applications benefit from true in-flight processing and immediate responsiveness, delivering measurable impact in areas where every millisecond counts.

However, this progress is uneven. Internal analytics functions—like finance, HR, and BI—continue to rely on traditional batch ELT workflows. Data for these domains is often landed into a warehouse and transformed on a fixed schedule, delaying insights and reinforcing silos between operational and analytical teams.

While pockets of real-time processing exist, governance, monitoring, and schema management vary widely across use cases. Stream ownership may be clear in one department but nonexistent in another. Tooling and standards are inconsistent, making it hard to scale real-time capabilities beyond isolated workflows. The result is an organization that's part real-time, part batch—with limited visibility and cohesion between the two worlds.

Level 4: Streaming-First Enterprise

In this mature state, organizations operate with a true streaming-first mindset. All data—regardless of domain—is treated as a first-class stream, and streaming becomes the default mode for both operational and analytical workloads. There's no distinction between "real-time" and "batch" use cases because everything runs on governed, low-latency pipelines designed for continuous processing.

At this level, enterprises implement a unified control plane, such as DeltaStream, to enforce consistent schema governance, end-to-end lineage, and freshness SLAs across the entire organization. Every domain—from finance and HR to marketing and product—consumes governed streams, enabling consistent, self-service access to reliable, real-time data. Dashboards, AI agents, operational triggers, and historical backfills are all powered from the same trusted streams. Teams no longer need to create special-case paths or duplicated pipelines. Batch is simply one mode of access among many, not the foundation of the architecture.

This level of maturity unlocks massive efficiency and innovation gains. It eliminates data silos, simplifies compliance, and ensures that every team—technical or otherwise—has access to trustworthy, always-fresh data to drive decision-making at the speed of business.

DeltaStream's Role in a Streaming-First World

Most enterprises today face a familiar dilemma: they need real-time insights to drive AI, BI, and operational decisions—but their current architectures are working against them. Analytics pipelines are fragmented, manual, and resource-intensive. Teams struggle with tool sprawl, infrastructure complexity, delayed insights, and runaway warehouse compute costs. These challenges stem from a common root: fragmented systems for streaming, real-time, and batch workloads.

DeltaStream Fusion is the first unified analytics platform that seamlessly combines streaming, real-time, and batch processing in a fully serverless environment. Designed for modern data teams, Fusion eliminates tool sprawl, overreliance on engineering teams, and infrastructure complexity that is costly to manage—enabling organizations to build powerful, real-time data applications and insights pipelines with ease and confidence.
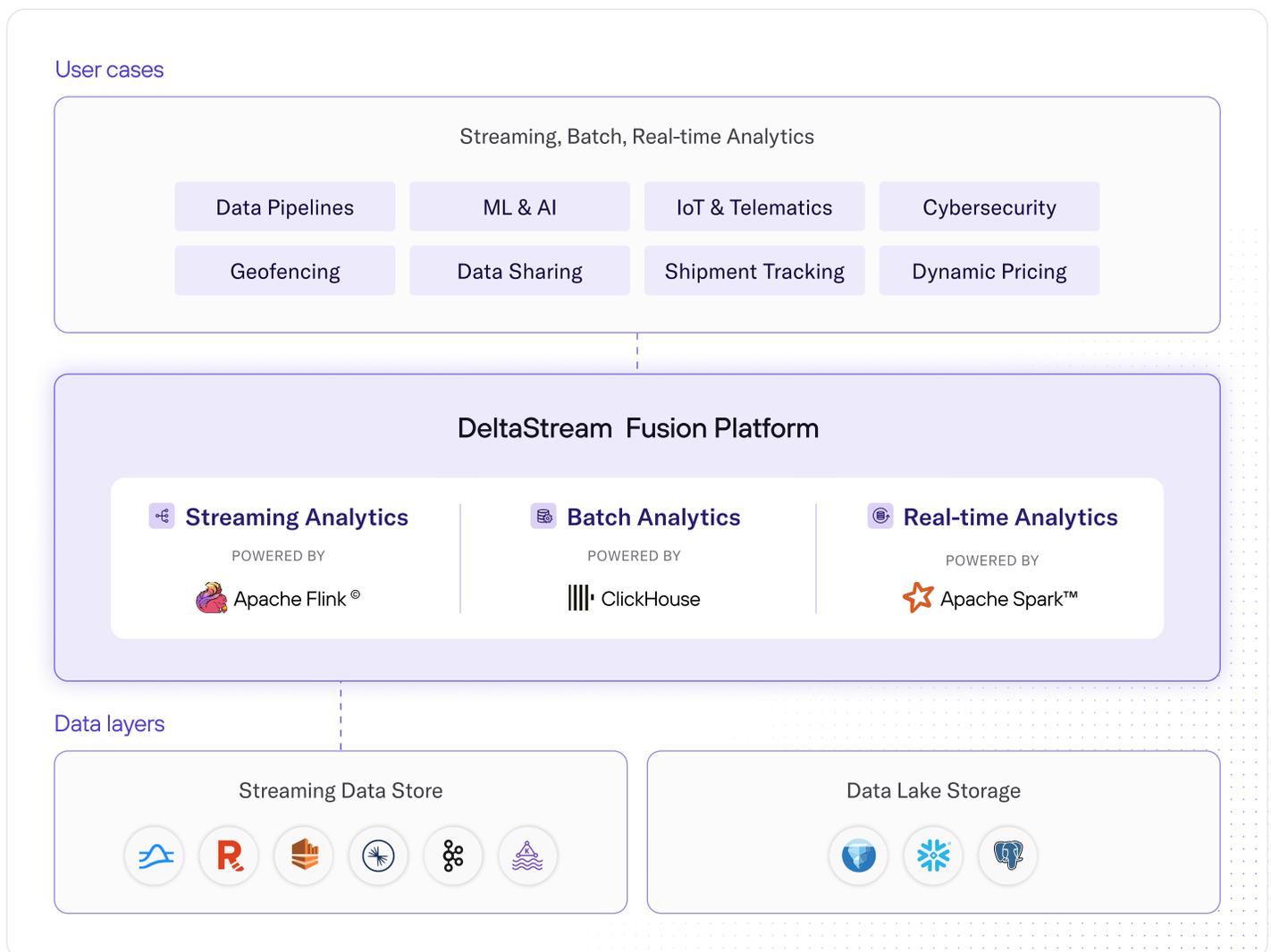
DeltaStream makes stream processing the default, not an afterthought—reframing analytics from a cost center into a real-time engine for business value. **DeltaStream solves this by:**

Unifying Analytics

DeltaStream Fusion eliminates fragmentation by consolidating batch, streaming, and real-time workloads into a single platform. Instead of stitching together Kafka, Spark, Airflow, and custom ETL frameworks, teams operate from one SQL-based interface that abstracts engine and infrastructure complexity. Fusion handles ingestion, transformation, and routing end-to-end—enabling seamless delivery to lakes, warehouses, or APIs. With Stream360 Observability built in, teams gain full transparency into lineage, transformations, and data flow, ensuring trust and operational clarity across the streaming estate.

Accelerating Time to Insight

Go from raw event streams to actionable insights in minutes—not months. DeltaStream automates infrastructure provisioning and workload execution so teams can focus on business logic, not orchestration. Just write in SQL—Fusion takes care of the rest, dynamically routing pipelines to the optimal engine. With Smart Engine Selection, Fusion evaluates the nature of each workload and dispatches it to Flink for streaming, ClickHouse for analytical queries, or Spark for batch jobs—maximizing speed while minimizing tuning effort.

Operating Intelligently at Scale

DeltaStream doesn't just run your pipelines—it optimizes them. The platform identifies changes at the event level and processes data incrementally, reducing compute costs and avoiding unnecessary refreshes. Pipelines are fault-tolerant, stateful, and elastic by default—ready to handle production-grade workloads. And by fusing real-time streams with historical context, teams can power advanced use cases like fraud detection, personalization, and anomaly detection with the full picture—not just recent events.

Its cost optimization impact is tangible. Traditional batch pipelines reprocess full datasets, burning warehouse compute. DeltaStream instead processes events incrementally, materializes continuous views, and keeps warehouse refreshes lean. The result? Real-time insights with half the infrastructure burden.
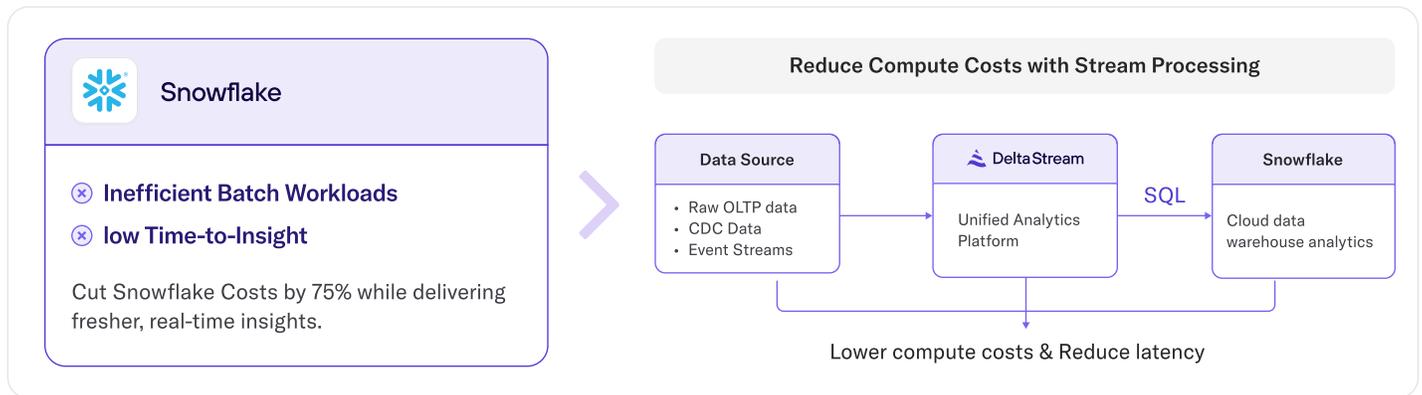
DeltaStream also powers the transition from batch to streaming without requiring a full replatform. Fusion connects directly to Kafka, Debezium, or CDC-enabled databases, processes events in-flight, and lands transformed data into Snowflake, Databricks, or lakehouses. Teams retain their existing destinations—just with fresher, more efficient data.

DeltaStream also elevates analytics into a core business driver. It turns insights into live data products embedded in apps, dashboards, and workflows—enabling fraud detection, personalization, anomaly alerts, and more. As companies shift from monthly reports to real-time decisions, DeltaStream delivers the control plane to make it possible.

From Cost Center to Value Engine

Keep your data active, responsive, and ready to create impact
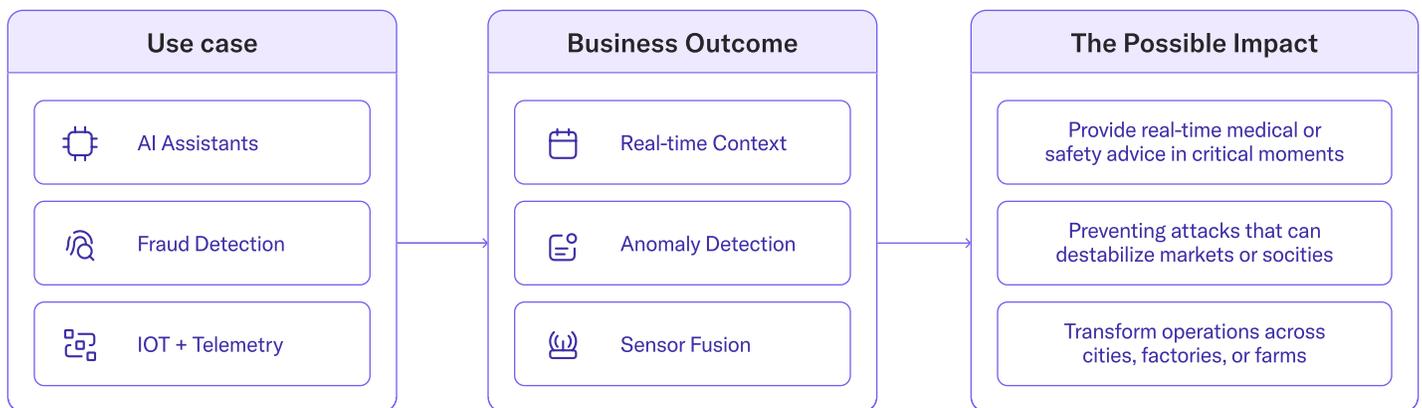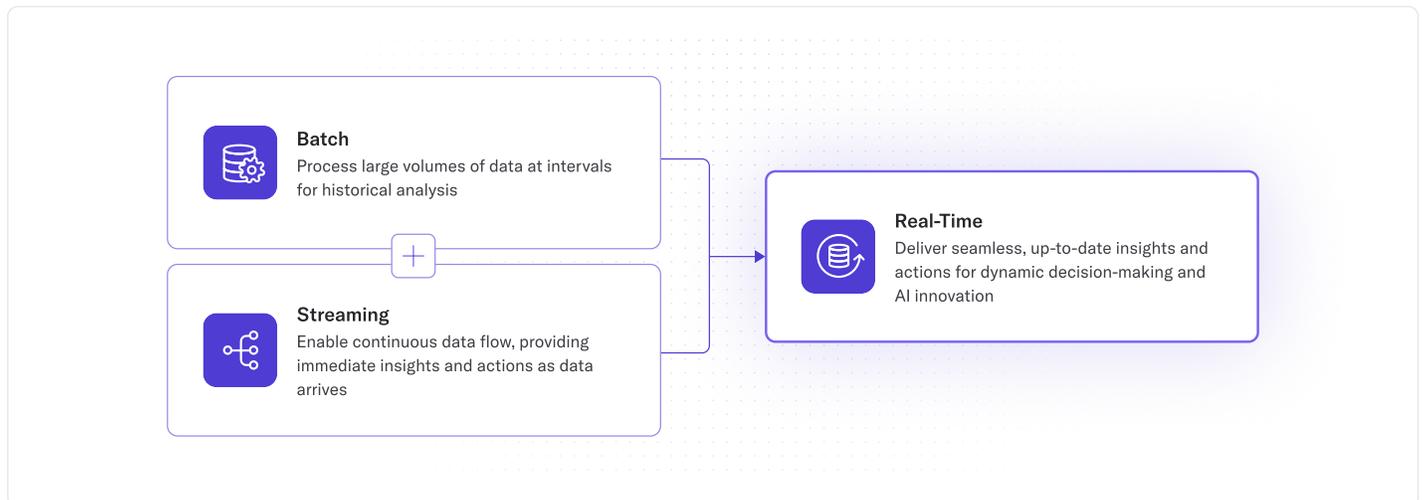

Whether you're consolidating ELT pipelines, slashing cloud costs, or enabling AI-driven applications, DeltaStream Fusion is the platform that unifies, simplifies, and accelerates your path to streaming-first.

Conclusion

Make streaming the foundation of your analytics with one unified, real-time platform.

The future of analytics is streaming-first, and it's closer than you think. With DeltaStream, you can stop batch and start making your data streams central to your analytics strategy.

One Platform. All Your Analytics. In Real Time. DeltaStream brings together streaming, batch, and real-time workloads under one unified, intelligent control plane. It gives your teams the tools to ingest, transform, and serve data at the speed of your business—no custom code, no duct tape, no compromise. Whether you're powering AI, BI, operational systems, or all of the above, DeltaStream ensures that analytics isn't just fast—it's foundational.

Batch
Process large volumes of data at intervals for historical analysis

+

Streaming
Enable continuous data flow, providing immediate insights and actions as data arrives

Real-Time
Deliver seamless, up-to-date insights and actions for dynamic decision-making and AI innovation

Book a complimentary 30-minute streaming architecture review with Hojjat Jafarpour, creator of ksqlDB and CEO of DeltaStream. He'll place you on the path to become a stream-first enterprise, and send a one-page action plan outlining:

- **Quick wins to cut latency or warehouse spend.**
- **Gaps to address before rolling out streaming ETL.**
- **Recommended architecture tailored to your tech stack.**

Ready to start? **[Schedule Your Review Now →]**

DeltaStream

DeltaStream

Power real-time decisions
and AI innovation with streaming-first infrastructure.