
# Building the Data Streaming Organization

## Driving Value From Data in the Enterprise

JON VINES

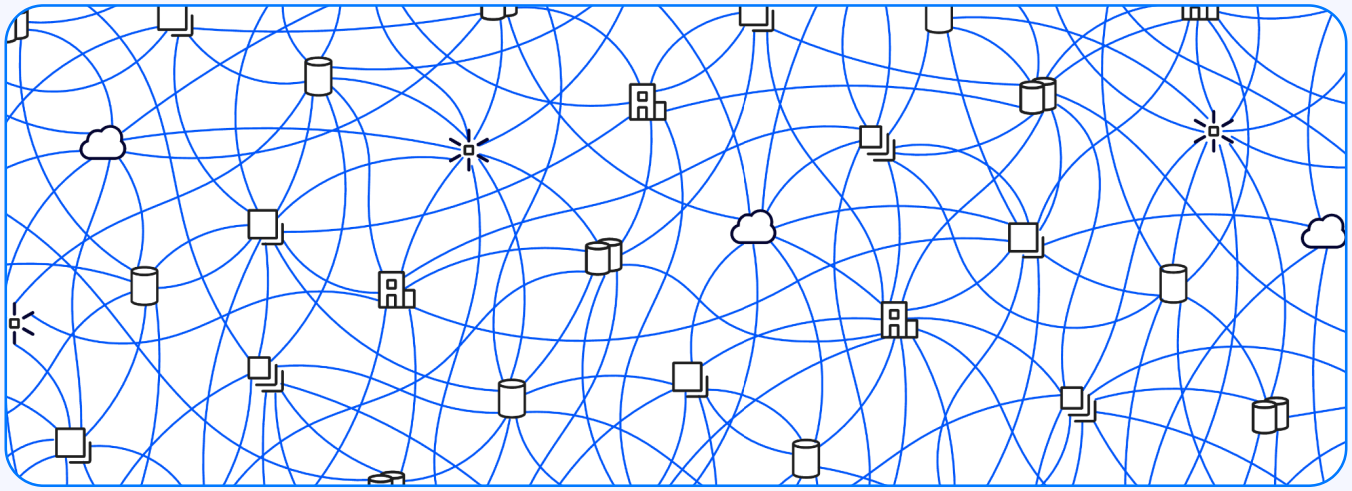
# Contents

# Building the Data Streaming Organization

## Driving Value From Data in the Enterprise

Organizations across all industries face a relentless demand for meeting increasingly complex customer needs while growing their businesses to meet revenue, growth, and regulatory targets. Access to quality data is critical as organizations strive to understand customers, track business performance, and ensure regulatory compliance.

We see this critical business data stored in many systems, which are connected by a spaghetti mess of complex and fragile sets of rigid point-to-point and batch connections—making it difficult, time-consuming, and cost-prohibitive to innovate. This data mess leads to several barriers that prevent organizations from uncovering the latent value locked in their current data approach.



Namely:

1. **It's challenging to discover, access, and use the data** that they currently have. We see this in the time it takes to uncover the data required to deliver new use cases. This relies on knowing the right people or falling back on anti-patterns such as replicating back from the data warehouse (since somehow, the data always finds a way to the data warehouse/lake).
2. **Businesses are siloed in their approach to technology, data, and processes.** Consider the myriad of technologies in play for integration and data access, the reach of analytical data processes pulling data into the analytics estate, and the rework required to prepare data for analytics, business intelligence, and data science. This siloed approach can result in much duplication of effort, with teams accessing the same data in various ways and building and operating similar technologies to facilitate data sharing.

3. **The available data is often outdated and inconsistent,** as evidenced by an organization's decision-making ability. We measure this in both the speed of insight and the speed of acting on the insight. Increasingly, customers expect organizations to respond in real time and to have a good understanding of their wants and needs.

> Read **Quick Thinking** to dive into some challenges that C-Suite executives face with real-time decision-making.

4. **Everybody is trying to solve the same problem of accessing timely data** and unlocking innovation through artificial intelligence (AI) initiatives. It's not enough to stand still and solve data problems as we've historically done. Customer expectations, technology innovations, and a need for timely access to data are driving organizations worldwide, in all industries and sectors, to tackle the data problem.

In answer to these challenges, we're witnessing the emergence of what we call the "Data Streaming Organization." Such an organization embraces the mindset of managing data as a product and creating intentionally shared, discoverable, quality, and timely data. It has formed the nexus of data-sharing strategies that span the moment when organizations create and curate data to facilitate real-time decision-making at the executive level. This new mindset manifests and presents itself as a network effect of shared data within the organization, powering the operations of businesses and the way customers interact with their products, as well as enabling innovation to meet future demands.

# Foundational Capabilities of a Data Streaming Organization

What compels an organization to change? And what are the tipping points that necessitate change happening now? We're in the midst of a rising set of conditions that fundamentally require us to shift our approach to how we manage data, including:

- Increasingly complex customer needs that require us to know customers and the context in which our organizations are operating intimately.
- Being agile enough to take advantage of technological advances, such as AI, that are having a profound impact on how our businesses operate.
- Having a strategic data platform that is scalable and cost-effective and that promotes managing data as a product for data sharing and reuse across the *whole* organization.

Technology is catching up rapidly to meet these conditions, if not meeting them already. Our organizational needs are fundamentally changing how we go about meeting them, too. Introducing the following foundational capabilities is critical in this journey:

1. **Treat data equally** across the whole organization by creating intentionally shared, discoverable, quality, and timely data managed as a product and made available for reuse across the entire organization.
2. Offer the technical capabilities to manage data as a product through a **unified self-serve internal developer platform.**
3. **Empower engineering teams** to process and govern data the moment they create it.

We describe these capabilities as the foundation of the Data Streaming Organization. We will not meet our customers' needs without introducing and fulfilling them.

# The Promise of Opening Up the Inherent Value of Data

Adopting the “data as a product” mindset at the organizational level, and applying it to data at the point of creation and through data streaming technologies, would significantly impact organizations’ ability to deliver new use cases to meet customer needs.

Organizations significantly benefit from managing and intentionally sharing data from the moment it’s created. The benefits include accelerating time to market, unlocking substantial return on investment (ROI), and driving cost reductions. Further, it readies organizations to be future-ready to meet the rise of technologies like generative AI, which rely heavily on accessible, high-quality data.

## Faster Time to Market

> *Companies that treat data as a product can reduce the time it takes to implement it in new use cases by as much as 90%.*
>
> — Harvard Business Review, “A Better Way to Put Your Data to Work”

Adopting the “data as a product” mindset significantly impacts the data discovery process when exploring new use cases across an organization. Several time thieves, such as unknown dependencies, unplanned work, and conflicting priorities[1] often arise due to a dependency on data from other business areas. These requests for data typically require the perceived owners of the data to make it available to the requestor, which is often unplanned. The mechanism for building the integration is generally ad hoc. When we can’t identify the owner, workarounds introducing anti-patterns emerge as the norm.

Imagine a world that compels engineering teams to create intentionally shared, discoverable, quality, and timely data at its source. Consider that a potential consumer of this data can request access and consume it in a self-serve manner and that it replicates automatically into the data lake/warehouse/lakehouse. This data-sharing approach helps us build an ecosystem that materially impacts time to market and value.

[1] “Making Work Visible,” Dominica De Grandis

# Unlock ROI

> *For a Fortune 100 company, just a 10% increase in data accessibility will result in more than $65 million additional net income.*
>
> — Forbes, "Better Data Quality Equals Higher Marketing ROI"

Identifying and creating this data doesn't come for free and requires a significant investment the first time we create it. These costs can compound further when we don't have the platform capabilities to facilitate intentionally shared, discoverable, quality, and timely data. However, once we've invested in our platform and in creating data in this way, we unlock a latent potential in how we can use it.

Consider the following properties of data:[1]

- Data can be held and consumed with massive concurrency, making relative unit cost negligible.
- Individual data sets cannot be substituted, leading to high valuation and leverage for data holders.
- Data's value is unknown until consumption, enabling inequitable value exchanges between parties.
- The value of data exponentially increases as it's accumulated, correlated, and contextualized.

Once the data is available and discoverable on the platform, it's available to all consumers (with the appropriate guardrails). Any subsequent data consumer adds to the data asset's inherent value, and the investment cost to create it becomes negligible.

# Drive Cost Reduction

> *Every year, poor quality data costs organizations an average of $12.9M.*
>
> — Gartner, "12 Actions to Improve Your Decision-Making"

We can approach the subject of cost reduction from many angles. Adopting Confluent and our Confluent Cloud capability can significantly decrease the total cost of ownership. The next layer of cost savings occurs when consolidating the variable flavors of Apache Kafka® deployment in the enterprise and building a technology capability that serves this need.

[1] "Unbundling the Enterprise," Matt McLarty and Stephen Fishman

The most significant and least understood cost saving is through the implication of treating data equally regardless of where we create it in the organization. Substantial time is spent in organizations that adopt a “replication of data” mindset—and it’s spent fixing quality issues, curating data many times, and discussing with many teams in the hope of understanding the data. However, adopting a “data as a product” mindset means we apply several characteristics that ensure quality and trustworthiness when we create new data. This approach to intentionally creating data means we can significantly reduce the time spent fixing and trying to understand the data we eventually access.

## Become Future-Ready

> *Our latest research estimates that generative AI could add the equivalent of $2.6 trillion to $4.4 trillion [in value to the global economy] annually across the 63 use cases we analyzed.*
>
> — McKinsey, “Economic Potential of Generative AI”

The demands on and promise of AI to organizations emphasize a growing need to create intentionally shared, discoverable, quality, and timely data. The following are the requirements for data supporting any AI strategy:

- A strong understanding of the value of data within the organization.
- Intentionally shared, discoverable, quality, timely data managed as a product.
- Transformational leadership to communicate and deliver AI-ready data and build AI capability.
- An AI (multi-agent) architecture with access to all pertinent real-time data and the ability to cross-communicate.
- An evolvable, future-proof data platform that reduces friction and cognitive load for its users.

Being AI-ready means being data-ready.

Read more about the **Business Value of the Data Streaming Platform** and a breakdown of how to start measuring the impact of the data streaming platform in the organization.

# The Data Streaming Readiness Framework

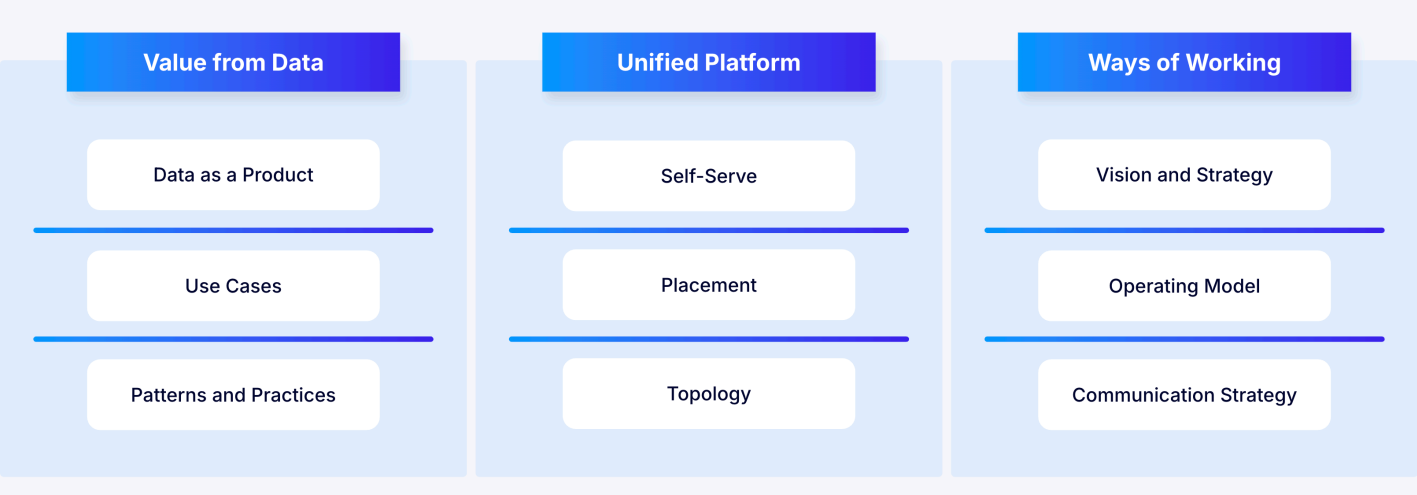



Figure 1: The Three Pillars and Nine Categories of the Enterprise Data Streaming Readiness Framework

Becoming a Data Streaming Organization requires thinking beyond the technical implementation of data in motion. It's a *socio-technical problem* that requires clear articulation and communication of how this approach brings **business value to the organization**, linking back to the goals it enables through agreed-upon measures of success. This articulation of value is essential in getting the proper executive backing to ensure organization-wide uptake and funding for the initiative while understanding the value of data.

Over the years, we've worked alongside our most mature customers, learning what has made them successful and what capabilities are critical to adopting data streaming across their organizations. We've refined this into a repeatable framework for becoming a Data Streaming Organization encompassing people, processes, and technology.

The Data Streaming Readiness Framework breaks down into three pillars:

1. Value From Data by Adopting a "Data as a Product" Mindset
2. A Unified, Self-Serve Offering of the Data Streaming Platform
3. Established Ways of Working and Guiding Vision and Strategy

These pillars provide a body of considerations and capabilities to assess a business that aspires to become a Data Streaming Organization. Each pillar is broken down further into sets of considerations at a finer level of detail. The framework provides us with a surface area for completeness to determine where a customer is and where it plans to be, and it helps build a roadmap for transformation.

# Value From Data

Creating and producing data within an organization is often a side effect of building software. The concerns through which teams across the business will use the data are usually localized to the business impact of the developed application only. This approach to creating data can lead to the challenges outlined earlier:

- It's challenging to discover, access, and use the data.
- Businesses are siloed in their approach to technology, data, and processes.
- The available data is often outdated and inconsistent.

When these symptoms are present in an organization, we deal with "just data." When we treat data as a byproduct of operational systems, data engineering teams are centralized and overburdened, and we see a disconnect for analytics teams from the sources and providers of data. Deriving value from data requires us to incentivize the organization to do something different: Treat data equally and start to manage it as a product.

## Treat Data as a Product

We define treating data as a product as building **intentionally shared, discoverable, quality, and timely data streams** made available throughout an organization. We can break down this definition into seven factors:

- **Valuable:** Data is a *value driver* within the organization. We can see a link between data use and business impact, including revenue, customer satisfaction, and processing time.
- **Manageable:** A data asset has a single, *clearly defined owner* aligned with the teams closest to the originating source.
- **Understandable:** *Documentation, metadata, lineage,* and additional insight are required to make the data easily understood and digestible.
- **Composable:** Data is created within the *guardrails that prioritize interoperability*, enabling the composition of different data assets for new use cases and ultimately driving data reuse.
- **Reliable:** Data availability, quality, freshness, delivery, and evolution are reliable. It has a *schema and associated data contract* clarifying every field's semantic meaning and acceptable values.
- **Measurable:** Data is measurable against a published set of service level objectives, so performance, value, reliability, and quality measures are easily obtainable and *build consumer trust*.
- **Available:** Data is *available at the pace required by the consumer*, which could be within milliseconds of the data being generated or within a window such as a minute, hour, or day.

We can compare “just data” to data managed as a product through these properties.

| Factor | Just Data | Data as a Product |
|---|---|---|
| Valuable | ✔ | ✔ |
| Manageable | ⊘ | ✔ |
| Understandable | ⊘ | ✔ |
| Composable | ⊘ | ✔ |
| Reliable | ⊘ | ✔ |
| Measurable | ⊘ | ✔ |
| Available | ✔ | ✔ |

Key: ✔ Yes, ✔ Maybe, ⊘ No

Need help deciding where to begin? Read **The Practical Guide to Building Data Products for IT Leaders** to quickly identify high-value data products that maximize ROI and demonstrate business value with a proof of value.

## Use Case Portfolio Management

Applying data streaming to real-world usage scenarios is a process an organization must consciously decide to pursue. Scenarios that could benefit from a real-time, event-driven approach must be identified and prioritized according to their business value. The organization’s architecture and development teams must be made aware of how to design event-driven applications, in contrast to alternative approaches such as batch or point-to-point messaging, and must practice how to think about the process through exercises such as event storming.[1] As data streaming adoption grows in the enterprise, effective portfolio management of existing and new use cases will help inform the evolution required for the platform.

Read **The Ultimate Data Streaming Guide** to learn how industries leverage streaming to power use cases that optimize operations and improve customer experiences.

## Patterns and Practices

As data streaming adoption grows, repeatable patterns and practices will emerge. These will help accelerate teams’ adoption of the paradigm and facilitate consistency across approaches. Enabling this consistency will also speed the process by which we realize value through adopting the platform, as what was once an exception is now “the way,” i.e., precedents set the standard, policies are formed or adapted, and processes become well-known.

[1] What is Event Storming?

# Unified Self-Serve Platform

To extract value from data quickly and easily, we need to build a unified self-serve platform that seamlessly integrates and enforces an organization's ways of working and governance requirements. A well-built and managed platform allows its tenants to get their work done without involving the central team that manages the service.

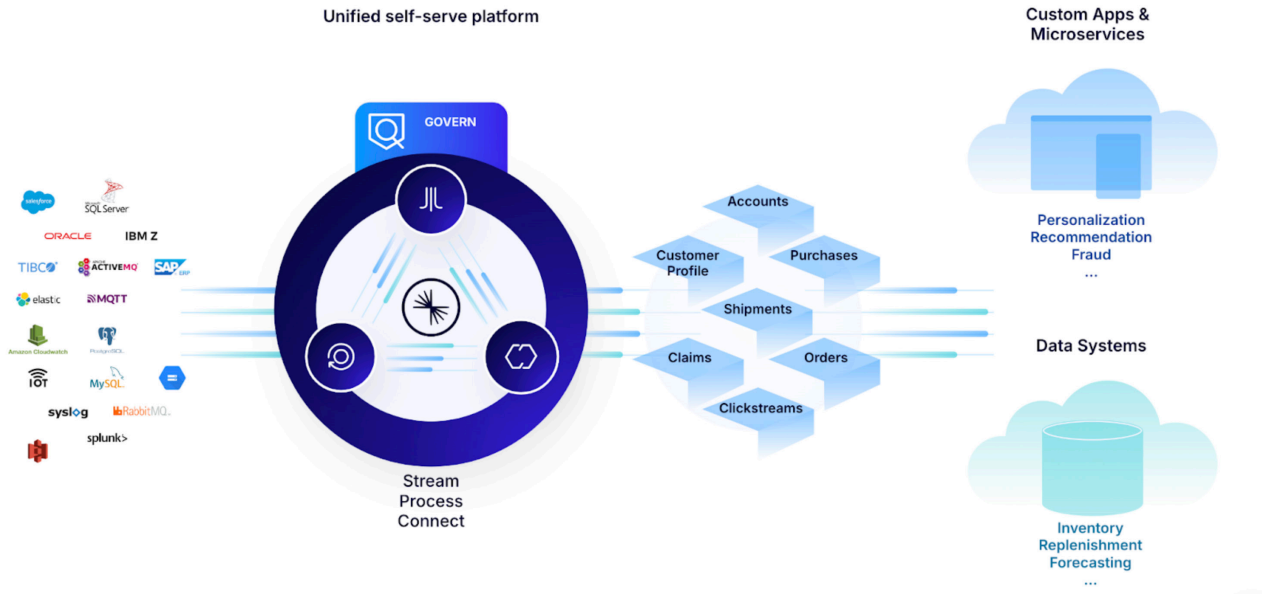



Figure 2: The Three Pillars and Nine Categories of the Enterprise Data Streaming Readiness Framework

The value realized through provisioning a self-serve data streaming platform that facilitates the delivery of data products has yet to be quantified. However, we can look to parallel industry trends to hypothesize the impact of automation and unification of the data technology landscape. In the book "Accelerate,"[1] Nicole Forsgren outlines the capabilities of an organization's ability to deliver software effectively (see Figure 3). Her research showed a direct correlation between continuous integration/continuous deployment (CI/CD) and test automation, among others, on the speed and safety aspects of software delivery. The role of the unified platform is CI/CD and automation for data, and we hypothesize a similar, if not greater, impact in the delivery of new and innovative use cases once we've created this virtuous circle of intentionally shared, discoverable data within an organization.

| | High Performers | Medium Performers | Low Performers |
|---|---|---|---|
| **Deployment Frequency** | On demand | Between once per week and once per month | Between once per week and once per month |
| **Lead Time for Changes** | Less than one hour | Between one week and one month | Between one week and one month |
| **Mean Time to Recovery** | Less than one hour | Less than one day | Between one day and one week |
| **Change Failure Rate** | 0%-15% | 0%-15% | 31%-45% |

Figure 3: Demonstrating the Impact of Automation on Software Delivery

[1] "Accelerate," Nicole Forsgren, Gene Kim, Jez Humble

## Self-Serve Internal Developer Platform

Internal developer platform capabilities must ensure that the data streaming platform integrates with the broader technology landscape and how the organization operates. The promise of decentralized data streaming requires that the platform uses self-service in a standardized, governed way that makes it easy to do the right thing. Failure to do so will lead to people drifting toward alternative solutions and shadow IT.

An internal developer platform includes features such as a developer portal, infrastructure as code, and integration with CI/CD pipelines. It also includes enablement materials such as documentation, code samples, and templates for getting started.

## Workload Placement

It's crucial to define and implement multi-tenancy guardrails to enable federated access to the platform and ensure that tenants can operate independently and with the security that their workload will run as expected. These guardrails include topic defaults such as naming, partition counts, and replication factors. Further, we will consider the governance and policies to be applied, where we place workloads, and where and how we access data on the cluster topology. We also provide the capability to monitor tenants and their adherence to predefined quotas.

## Architecture Topology, Deployment, and Automation

A unified data streaming platform must be deployed across multiple environments (development, test, production, etc.), considering an organization's throughput, latency, networking, security, auditing, and monitoring requirements. A platform may comprise many federated clusters owned and operated by the data streaming function. It requires automation for maintenance, upgrade, and replication.

We must define a data streaming platform cluster strategy to address high availability, disaster recovery, multi-geography, hybrid and multi-cloud, multi-tenancy, workload placement, and data federation. Clustering considerations include all workload types that run on Confluent, including topics, connectors, stream processing tasks, and schemas. We need to support this strategy with robust recovery processes.

# Ways of Working

Embedding data streaming as an organizational capability requires transformational leadership and organizational awareness. Clearly articulating the roles and responsibilities plays a significant role in the success of any transformation. How will required changes impact the rest of the organization, what organizational policies must be complied with, and what processes do we need for users to use the platform efficiently? Identifying and agreeing on how we implement the roles, responsibilities, and processes we require can help determine the operating model needed to execute the strategy.

> *[T]he organizational structure must coordinate accountabilities to support the goals of delivering high-quality, impactful [data].*
>
> — "Thinking Environments," Schwartz et al.[1]

## Vision and Strategy

We need to set a clear vision[2] for how we will use data streaming to share data, and it should define the technology's unique position in the enterprise technology landscape compared to other technologies (integration platforms, messaging, etc.). We should create a prioritized roadmap for adoption, and it must address identifying capability gaps in awareness, operating model, talent, capacity, data, and technology. The execution of this roadmap must initially focus on addressing the capability gaps necessary to form the foundation for early delivery of value, building credibility for the transformation.[3]

Fundamental to any strategic initiative is measuring success by establishing and agreeing on key performance indicators and objectives and key results. This enables both course correction and quantification of success. Articulating our measures of success in this way will allow the visibility of ROI and assess alignment with organizational strategic objectives.

## Operating Model


Figure 4: There are many ways to represent who is responsible for ensuring the success of the data streaming platform, including a combination of all three models.

A model that the data streaming function will implement through the platform must also be defined. It should create a set of standard value streams (application onboarding, accessing existing data, decommissioning workloads, etc.), teams and responsibilities, interaction patterns, maintenance and upgrade policies, usage policies, chargeback handling, and incident management and communication.

We can address organizational risks through governance, policies, and standards. A governance process will involve managing a partnership between expert groups within the organization (information security, data governance teams, etc.) and the group responsible for data streaming, which will take this input as requirements to implement within the platform through procedures and controls. The group will report on its success through measurement.

[1] Thinking Environments, Mark Schwartz, Jason Cox, Jonathan Snyder, Mark Rendell, Chivas Nambiar, Mustafa Kapadia, 2016
[2] Moving Up the Curve: 5 Tips For Enabling Enterprise-Wide Data Streaming, Lyndon Hedderle
[3] Defining your 'true north': A road map to successful transformation, Kevin Carmody et. al, May 2024

## Communication, Positioning, and Community Building

Community building is crucial in building awareness and leadership buy-in. This community serves as a hub for exchanging use cases and successes. A vibrant community educates, inspires, and ensures the platform's continued relevance and usage. Those managing the platform should also be responsible for enablement, migration, education, and continuous inspiration.

Internal positioning, branding, and a strong communication strategy are crucial to the success of platform adoption. When discussing the advantages of data streaming technology, we can't rely on one common understanding for everybody. That's why it's essential to employ different tonality, messaging, and forms of presentation for each persona, each level of the maturity curve of a department and person, and their learning types. Through ongoing evangelism of success and value, the broader organization can attract new tenants to reuse existing data while bringing on new workloads and use cases, enhancing the platform's value through a flywheel effect.

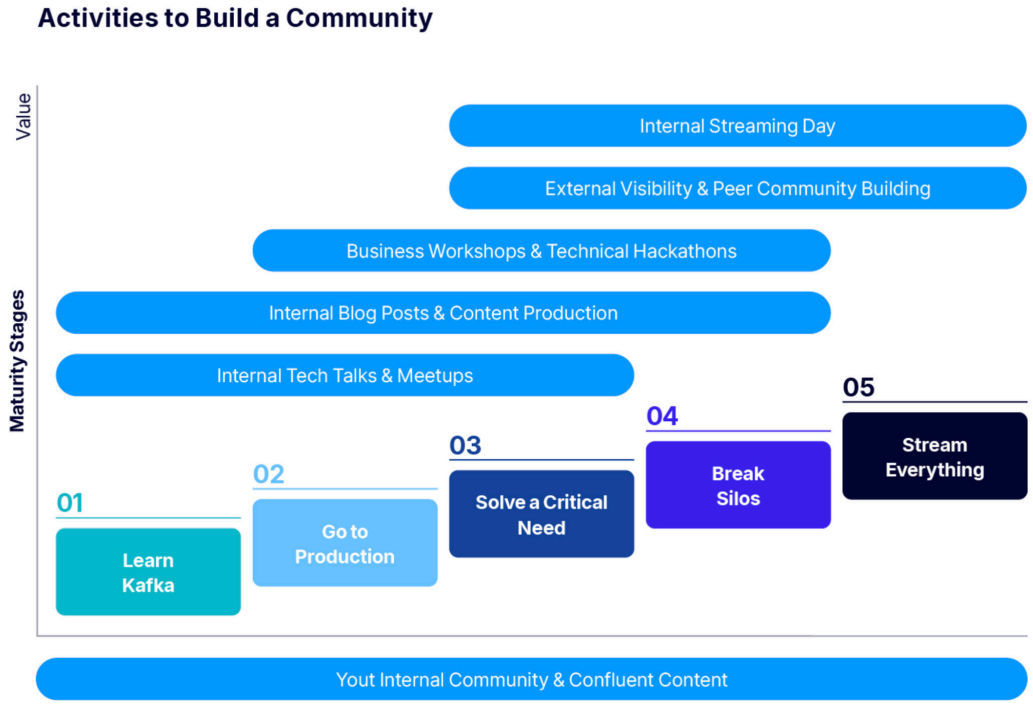



Figure 5: The Stages of Evolution as the Data Streaming Community Matures

A foundational part of the communication strategy will be gathering, curating, and presenting the developed patterns and practices to the organization across all available channels. Subject areas can include deployment architectures, producer and consumer considerations, how to design and build event-driven applications, considerations for outage handling, data integration approaches, monitoring, benchmarking, testing, resiliency patterns, and business value–generating use cases and success stories.

Read Chapter 4 of **The Ultimate Data Streaming Guide** to learn more about building a community and communication strategy based on real-life examples from BMW Group, Michelin, and many more.

# Wrap-Up

Confluent's data streaming platform is at the core of organizations aspiring to become Data Streaming Organizations. It enables connecting data from any system, governing to ensure data quality, and processing data for appropriate context—all in real time through streaming. The data streaming platform comprises the minimum technical capabilities required to deliver streaming data, managed as a product, in an organization.

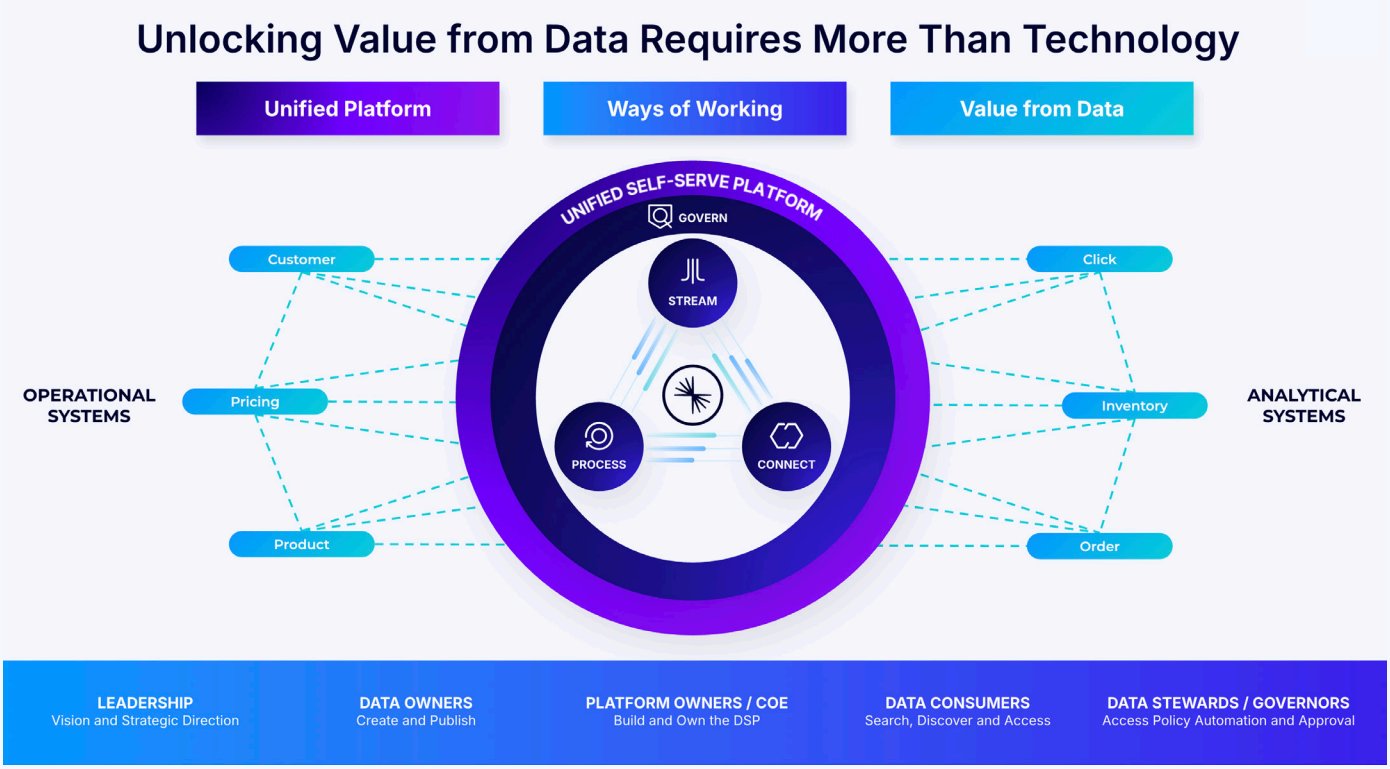



Figure 6: Overlay of the Three Pillars of the Data Streaming Organization and Their Place Alongside the Data Streaming Platform

We propose a framework describing the people, process, and technology capabilities that organizations require to deliver intentionally shared, discoverable, quality, and timely data across an enterprise. Providing this capability requires organizations to generate, process, and respond to data when the action occurs, extracting the maximum value from the data as it's generated. To facilitate data creation in this manner, organizations need a clearly defined technology and data-sharing strategy informed by the value of data that we can identify and realize. To produce data in the way that organizations need, we require a unified approach to our data streaming platform and the execution of our strategy through agreed-upon ways of working, defining ownership for delivery of the platform and the data we deliver through it.

These data streaming capabilities will become foundational as our organizations evolve and respond to the changing demands described in this paper. Within five years, we'll celebrate organizations that successfully solve the data-sharing problem through technologies such as the data streaming platform. This success will reflect in outstanding annual results, stellar customer satisfaction, and recognition of the emergence of the Data Streaming Organization.

# Getting Started

Getting help on your journey to becoming a Data Streaming Organization is closer than you think. Confluent has dedicated resources to help you at this critical inflection point in your efforts to build and scale your Data Streaming Organization capabilities.

If your business needs assistance transforming its Data Streaming Organization capabilities, contact your Confluent account team or a **Confluent Executive Advisor** who can help you discover the best way to become one. This assistance may include any of the following:

- Discover where your organization is with our Data Streaming Readiness Assessment.
- Deep dive into the capabilities required and where to take the next best action with an Activation Workshop.
- Identify high-value data assets and get the whole organization on board through Event Storming.
- Engage with our Professional Services team and Partners, home to world-class data streaming platform experts, to design and build a platform capable of establishing a Data Streaming Organization.