# Precisely Data Quality for Kafka

As the volume and velocity of data continues to grow, streaming data represents a paradigm shift that introduces new data quality challenges.

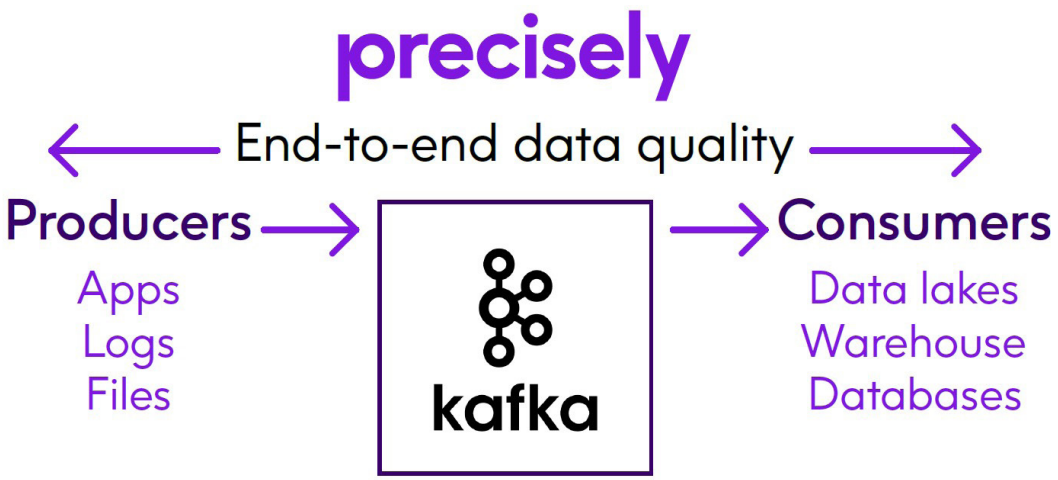
## The Challenge

The speed of business today demands that organizations enable access to real-time data. This has given rise to event-driven architectures. This fundamental shift from batch processing to streaming data provides new opportunities for increased data delivery reliability, more nimble reactivity and faster business insights. But it also presents new risks to data integrity as streaming sources, data volumes and architectural complexity continue to grow, affecting data quality Among the many streaming data options, distributed streaming platform Apache Kafka has become a top choice for real-time data communication. Kafka is an agile, high-throughput, low latency option for managing data in motion, but it cannot ensure the reliability and accuracy of real-time data streams The options that exist to ensure data quality of data in Kafka streams is often limited to very technical options that require a robust skill set and are limited in their ability to apply business rules to the data. Business teams expect the same logic to be applied to real-time data that would be applied in a batch mode.

## The Solution

As reliance on streaming data continues to grow, so do demands from data consumers who expect real-time intelligence that's complete, consistent and trustworthy. That's why organizations need to ensure end-to-end data quality in streaming environments, as messages are sent from the sources that create them (producers) to the sources that consume them (consumers).

Businesses need a solution that provides a wide range of data quality capabilities, from in-line checks and validations to batch processing for robust reconciliations. Precisely's data quality solutions for streaming data give organizations confidence that the quality of streaming data won't degrade in motion and can immediately be acted upon in order to build trust throughout the business.



## Precisely Data Quality for Kafka delivers:

**Trust in streaming data with producer-to-consumer validation**
Reduce enterprise risk by ensuring that streaming data is validated, reconciled, and timely, to produce meaningful and reliable data insights.

**Accelerated time-to-value**
Automatically connect and configure Kafka stream schema with zero-code setup based on integration with your platform schema registry. Easy setup combined with drag-and-drop rule development shortens time to value.

**Clarity and insight across your data pipeline**
Validate data for streaming and non-streaming data producers to allow for validation across the entire data pipeline, from raw data in the source system to Kafka message, to consumer (database, file, application, etc.).

**Single solution reduces support and costs**
Provide an integrated toolset in a single solution, reducing the costs of multiple tools and the number of supported applications. Experience a breadth of functionality that helps consolidate required resources, skill sets and training with a single solution for IT, business users and audit.

## Key Solution Features

**Support for real-time or batch validation on streaming data**
Easily build user-configurable rules and apply them to real-time or batch streaming data to support inline or time-windowed validation . Behind the scenes these rules are , powered by the latest Spark Structured Streaming and SparkSQL.

**Apply validations and business rules**
In addition to applying data validations such as completeness and conformance to data types, apply complex business rules like ensuring the product code is valid for the line of business. You can even enrich data within the process. An example is to add a product name based on the product ID.

**Ingest, enrich, transform and output different data types**
A wide range of data type options can be captured as input and written as output, allowing users to input as one data type and write out as another. This flexibility allows traditional data sources like databases and files to be enriched or transformed and then output as Kafka messages, or vice versa. It also allows the blending of existing architectures without waiting for all data to fully transition to streaming.

**Apply machine learning to identify streaming data patterns or anomalies**
Discover patterns, highlight outliers and apply analytics across multiple facets of streaming data using embedded machine learning capabilities.

**Manage exceptions and visualize results**
Identify duplicate or failed messages and route them via workflow to be assigned, managed and resolved. Visualize the results of your streaming data validations to gain deeper business insights while sharing and collaborating with teams through dynamic dashboards.