

# Ingest and Process Streaming and IoT Data for Real-Time Analytics

## Benefits

- Gain real-time operational intelligence with streaming analytics
- Increase developer productivity with an easy-to-use design interface
- Deliver information at any latency with a unified approach
- Adapt quickly to new streaming data between multiple sources and targets
- Simplify configuration, deployment, administration, and monitoring of real-time streaming

## Informatica Enterprise Streaming and Ingestion Provides End-to-End Multi-Latency Data Management With Versatile Connectivity

### Streaming Data Collection and Ingestion Requires a New Approach

In today's data-driven economy, organizations are using streaming, IoT, and real-time data to unlock new business opportunities and gain competitive advantage. Streaming data enables real-time, data-driven decision making and improves operational efficiency with initiatives such as predictive maintenance, fraud detection, and enhanced customer experience with real-time offers.

Collecting and managing streaming data requires a new approach in conjunction with traditional, batch-oriented data collection processes. Batch solutions that require intermediate steps (such as storing the data) are not designed to ingest data in real time and require close monitoring to prevent log jams. Nevertheless, an existing batch pipeline can be leveraged for streaming data collection and management without requiring investment in completely new plumbing for streaming data.

### High-Performance Edge Data Processing Needs Filtering and Management of Data Drift

Streaming data needs to be collected in real time and moved onto a data lake with the ability to manage peak and unusual traffic. The data volume may be very large, so applications need to filter out some of the non-critical data before ingesting the data into the lake.

The data collected from streaming and IoT systems can change over time due to various events such as firmware upgrades or changes to the source systems. This is referred to as data drift. It is important for the streaming solution to automatically address the drift in the data without interrupting the data management process.

### Data Integration Needs to Support Multi-Latency Data Management

The velocity and the volume of streaming data makes it challenging to integrate and transform streaming data in real time. A simultaneous dual-mode approach is necessary to support processing of real-time streaming data as well as batch data. Resource constraints during peak hours may hinder performance of stream processing, requiring organizations to move some of the processing to batch. Developers need to be able to design data streams once and deploy them for production on thousands of IoT devices, allowing the system to manage the data processing and meet service-level agreements.

### Data Needs to Be Collected from a Variety of Sources

In the world of streaming data there are many standards and a variety of protocols (e.g., MQTT, OPC, and AMQP) as well as a variety of data types (JSON, XML, Avro). In addition, data comes in structured, semi-structured, and unstructured formats. Data also exists in files, databases, and applications. That data needs to be collected as part of the overall streaming solution. Streaming data also needs to be collected from messaging and database environments to capture changes as they are triggered based on business events.

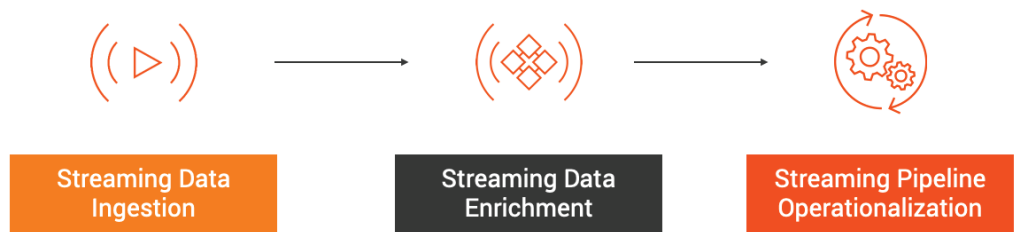


Figure 1: The streaming data journey.

### Key Benefits

#### Drive More Value Faster from Real-Time Streaming Initiatives

Enable real-time operational intelligence with a single streaming analytics solution that can sense, reason, and act on the streaming data. The solution captures change data as well as real-time data and provides data management to filter, transform, aggregate, enrich, and process the data. Last, the data is delivered to the data lake for analytics and to feed real-time actions.

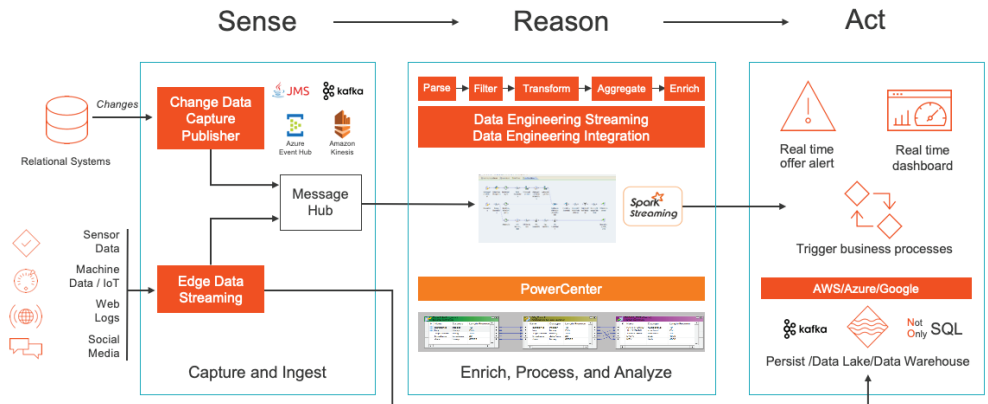


Figure 2: Real-time streaming data management.

### **Gain Real-Time Operational Intelligence**

By collecting and integrating streaming data, companies achieve a truly holistic real-time picture of business activities. Informatica® Edge Data Streaming is designed for continuous ingestion and timely processing of streaming data on commodity hardware with no need for intermediate data staging or brokered mediation. Adding new data sources and processing targets is fast and easy, and the data flow between sources and targets is simplified so your IT team can spend less time managing your streaming data and more time finding ways to leverage it for strategic gain and operational efficiency. Edge Data Streaming also enables edge transformations so that low-quality data doesn't make it into the data lake and so avoids turning into a data swamp.

Informatica Data Engineering Streaming helps you transform data into intelligence faster so you can identify patterns, detect abnormalities, and alert decision-makers to events that indicate imminent risk or opportunity—all while storing the data in processing environments like Hadoop for ongoing use and correlating it with historical information for predictive analytics.

### **Future-Proof Your Investment With a Unified, Multi-Latency Approach**

Optimize your stream and batch data processing based on available system resources and business service-level agreements. Data processing can range from sub-second stream processing on Spark Streaming, to batch processing on Hadoop at scale, without having to redesign or rebuild data pipelines. You can build data pipelines once and run them at any latency without needing any specialized development.

### **Support a Wide Variety of Sources and Targets**

Informatica Edge Data Streaming includes lightweight agents that provide out-of-the-box support for a wide and growing number of streaming data sources and targets. These agents minimize the need to develop source and target adapters internally, speeding the process of integrating streaming data into processing environments. The agents are automatically deployed based on a user-defined topology configuration and directly connected via the high-performance message bus. This allows Edge Data Streaming to avoid data staging, moving data instead with simple, one-hop flows from sources to targets. At the source, agents read and dispatch data even as it is logged to a file. At the target, agents receive and write that data to the appropriate processing environments.

Informatica Change Data Capture recognizes business events—such as customer creation or order shipment—and captures database inserts, updates, and deletes as and when they occur. The database change data can be published directly onto the messaging hub without intermediate staging tables.

## About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category, or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities, or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

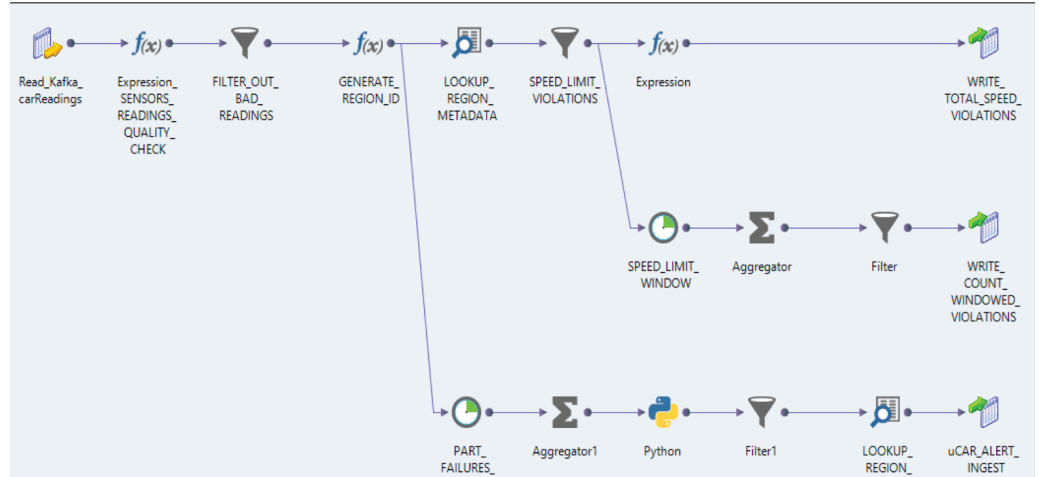


Figure 3: The data pipeline workflow development environment.

## Increase Developer Productivity With an Easy-to-Use Design Interface

Time-to-value measures how quickly you can progress from design, build, and test to deploy and maintain. Informatica Data Engineering Streaming increases development productivity up to five times over hand coding. Using a visual development environment and prebuilt dynamic templates, developers can build data streams without specialized knowledge of Spark Streaming concepts or languages and rapidly deploy data streams into production with simple configuration parameters. This level of abstraction between the visual development environment and the underlying processing engine enables you to deploy data streams anywhere, whether on-premises or in the cloud.

## Next Steps

Learn more by visiting the [Informatica Data Engineering Streaming](#) and [Data Engineering Integration](#) product pages.



Worldwide Headquarters 2100 Seaport Blvd., Redwood City, CA 94063, USA Phone: 650.385.5000, Toll-free in the US: 1.800.653.3871

IN17\_0919\_03714