Informatica Data Engineering Streaming

Benefits

- Sense, reason, and act on trusted real-time streaming data

- Enable real-time streaming analytics use cases with improved streaming data quality

- Accelerate your journey to the cloud with best-in-class streaming connectivity

- Run mission-critical streaming applications with enhanced operational efficiency

- Enable real-time operational intelligence with data engineering streaming analytics

- Reduce time-to-value with increased developer productivity

- Deliver information at any latency with a unified approach

- Simplify configuration, deployment, administration, and monitoring of real-time streaming

Gain Trusted Insights for Real-Time Analytics

Businesses today have an unprecedented opportunity to gain insight from a steady stream of real-time data—for example, transactions from databases, clickstreams from web servers, application and infrastructure log data, geolocation data, and data coming from sensors or agents placed on the almost endless variety of devices and machines making up the Internet of Things.

This continuous flow of messages and events can increase the effectiveness, agility, and responsiveness of decision-making and operational intelligence. However, as data flows in at high rates, it accumulates quickly into large volumes. Organizations can derive maximum value from data only if they can gather and analyze it immediately and at an ever-increasing scale.

Modern Scalable Architecture for Streaming Analytics

Informatica® Data Engineering Streaming allows organizations to prepare and process streams of data and uncover insights while acting in real time to suit business needs. It can scale out horizontally and vertically to handle petabytes of data while honoring business service level agreements (SLAs).

Informatica's approach to real-time data ingestion and management starts with collecting the raw data from various sources and ingesting the data into a data lake or messaging hub. Informatica also offers data transformation and data enrichment capabilities to process the streaming data and make it available for operationalization and downstream analytics.

Streaming Data Ingestion → Streaming Data Enrichment → Streaming Pipeline Operationalization

Figure 1: The streaming data journey—ingestion, enrichment, and pipeline operationalization.

Informatica Data Engineering Streaming provides prebuilt, high-performance connectors such as Kafka, HDFS, Amazon Kinesis, NoSQL databases, and enterprise messaging systems and data transformations to enable a code-free method of defining your data integration logic. Productivity and maintenance are dramatically improved by the automatic generation of whole classes of data flows at runtime based on design patterns.

Low-Latency Data Architecture Built to Scale With Evolving Cloud and Open Source Technologies

Informatica Data Engineering Streaming is built on best-in-class open source technologies in an easy-to-use, enterprise-grade offering. It primarily uses open source Spark Streaming under the covers for stream processing and supports other technologies like Apache Kafka, Azure Databricks, and Databricks Delta. As new technologies inevitably evolve, Informatica Data Engineering Streaming adapts, using the same data flows so you don't have to rebuild them. And you can schedule data flows to run at any latency (real time or batch) based on the resources available and business SLAs.

Informatica offers the Sense-Reason-Act framework for real-time data ingestion. The framework provides end-to-end data engineering capabilities to ingest real-time data, apply enrichments on the data in real time or in batches, and operationalize the actions on the data in a single platform using a simple and unified user experience.
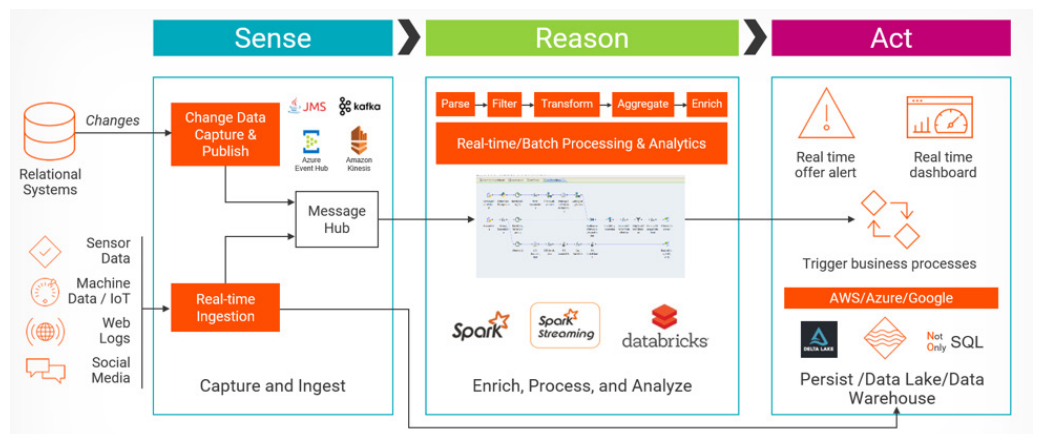

Figure 2: The Sense-Reason-Act framework.

Key Features

High-Performance Streaming Analytics With Reliable Quality of Services
Collect, transform, and join data from a variety of sources, scaling for billions of events with a processing latency of less than a second. You can store data in a data lake for ongoing use and correlate streaming data with historical information. Choose from several qualities of service levels according to your business requirements.

Real-Time Processing with Business Rules
Write and execute a set of event-driven business rules against transformed and enriched streams of data through an easy-to-use intuitive rule builder. Users can define patterns, abnormalities, and events that, should they pose imminent risk or opportunity, trigger alerts so the right people can respond in real time.

Faster Stream Data Management
Develop streaming processes faster with an extensive library of prebuilt transforms running natively on Spark Streaming to process all types of data at scale. In addition to running on Spark Streaming, Informatica Data Engineering Streaming uses secured Kafka (with Kerberos) as the data transport across mappings and data replay for recoverability; HDFS as a highly-available persistence store for recoverability data; and speedy in-memory capabilities to avoid continuous database lookups.

Unified Low-Latency Approach
Ensure speed and flexibility with a single, consistent data-processing approach for all latencies. Developers design data streams once and deploy them once. Existing data pipelines are easier to maintain and face less risk as Spark Streaming evolves, or if a new stream-processing engine is adopted. As a result, data streams and new innovations are implemented faster with less impact and risk to production systems.

Stream Processing for Virtually All Types of Data
In the world of fast data there are many different data formats and types produced by machines and IoT devices. Informatica Data Engineering Streaming processes all types of data including complex hierarchical data objects in a variety of formats (e.g., JSON, XML, Avro, CSV) and types (e.g., Array, Struct, Record and Maps, Nested HTYPE).

Spark Structured Streaming
Process streaming data based on event time instead of processing time with support for Spark structured streaming. Informatica Data Engineering Streaming also supports streaming-specific capabilities such as "out of ordered delivery of streaming data" with watermarking.

Cloud-Ready Streaming
Easily develop both batch and streaming pipelines with support for Databricks clusters in Informatica Data Engineering Streaming. Customers can now run streaming jobs on Azure Databricks clusters with Databricks Delta as the target.

Simple, Centralized Configuration, Administration, and Monitoring
Informatica Data Engineering Streaming is built on the Informatica Intelligent Data Platform™. Its administrator tool lets you easily manage and monitor your system, users, and deployed mappings.

High Availability, Scalability, and Architectural Flexibility

Informatica Data Engineering Streaming supports high availability, automated failover configuration on commodity hardware (with no need for a shared file system), and guaranteed delivery of data. This is required for uninterrupted processing of streaming data, to ensure data is never lost and SLAs are met. Increasing horizontal and vertical scalability is as easy as deploying more Spark nodes. The flexible architecture supports changing business requirements, with sources and targets connected in any pattern.

Advanced Streaming Data Transformations

The need to utilize data analytics on fast-moving streaming data for improved results is critical to a business's success. Informatica Data Engineering Streaming can apply data quality transformations on streaming data that help drive real-time use cases such as targeted marketing campaigns, predictive maintenance, fraud detection, and clinical research optimization. Customers can now be certain of the quality of streaming data loaded into their data lakes. Informatica Data Engineering Streaming supports four transformations: Classifier, Standardizer, Parsing, and Address Validation.

Intelligent Stream Data Parsing

Run mission-critical streaming applications with enhanced operational efficiency using Confluent Schema registry. Automatically parse Avro messages in Kafka using Confluent Schema and complex streaming data with intelligent structure discovery powered by the CLAIRE® engine. Easily handle schema drift and evolving schema.

Enhanced Connectivity Across AWS and Microsoft Azure

Informatica Data Engineering Streaming enables ingestion and processing of real-time streaming data into Amazon S3 and Azure ADLS Gen2, to accelerate your journey to cloud. It fully supports Amazon Kinesis Streams as a source, Amazon Kinesis Firehose as a target, and Amazon EMR in streaming mode, making it easy to collect, deliver, and process large amounts of real-time data efficiently.
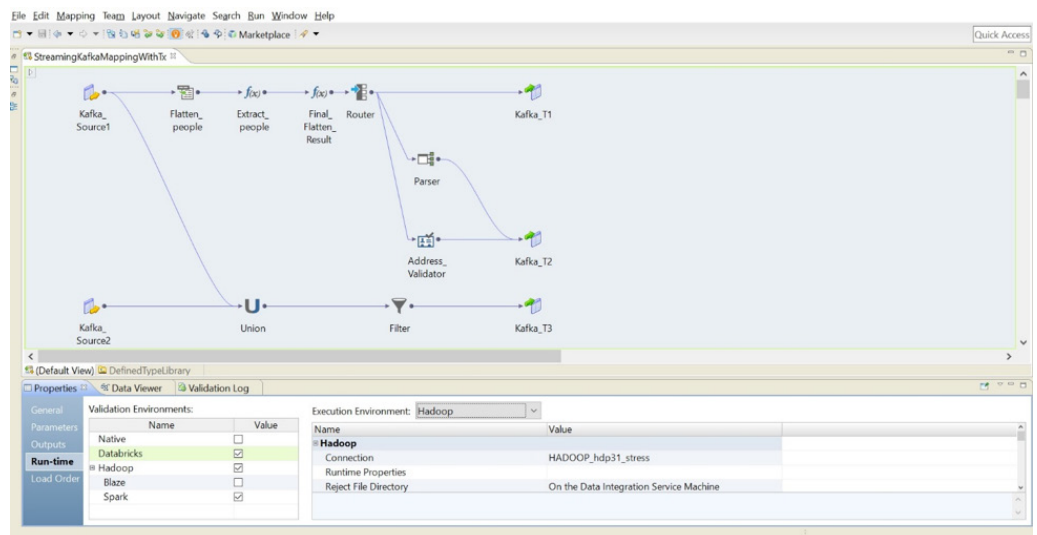

Figure 3: The Informatica Data Engineering Streaming visual development environment provides up to five times the productivity of hand coding.

Key Benefits

Get More Value out of Real-Time Streaming Initiatives
Enable real-time operational intelligence with a single streaming analytics solution that can capture, transport, refine, enrich, process, and distribute streaming data in real time. Combine real-time data from sensors, devices, and machine logs with other enterprise data such as transaction, customer, product, and reference data to discover and respond to actionable insights at the speed of business.

Future-Proof Your Investment With a Unified Low-Latency Approach
Optimize your stream and batch processing based on available system resources and business SLAs. Data processing can range from subsecond stream processing on Spark Streaming, to batch processing on Hadoop, 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.

Reduce Time-to-Value With Rapid Development
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 and 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.

Minimize Risks Associated With Complex and Evolving Open Source Technologies
Informatica Data Engineering Streaming minimizes risks associated with rapidly evolving technologies such as Spark and Spark Streaming. The IT organization can make one investment that continues to work with the changing technology landscape, providing a single, consistent data processing approach for all types of data at all latencies. Data pipelines are easier to maintain as emerging technologies continue to evolve and change, which means your development is future-proof to quickly adopt the latest innovations in real-time streaming.

Troubleshoot Issues More Effectively for Improved Customer Service
Most companies get feedback from customers, directly and indirectly, in a few ways: call centers, emails, support tickets, customer service bots, server logs, and more. With Informatica Data Engineering Streaming, companies can perform real-time correlation of data that might have otherwise been disconnected—for example, a software-as-a-service (SaaS) application provider could correlate server logs with questions to customer service bots to discover, at a faster pace than previous methods, that a service has crashed.

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.

Transform Business Insights

Informatica Data Engineering Streaming generates significant business value for IoT applications. Here are some of the industry use cases where streaming analytics is helping enterprises drive competitive advantage:

- **Preventative maintenance** - Real-time streaming analytics can reduce operational and equipment costs by minimizing unplanned outages and avoidable site and maintenance visits.
- **Retail** - Real-time inventory updates help drive business processes for inventory and pricing optimization, as well as optimization of the supply chain, logistics, and just-in-time delivery.
- **Smart energy** - Real-time monitoring of smart meters permits smart pricing models for electricity, as well as integration with renewable energy generators to optimize power generation and distribution.
- **Industrial automation** - Streaming and predictive analytics enable manufacturers to optimize production processes and product quality, including automated alerts and production shutdowns when quality levels are breached.
- **Healthcare** - Real-time data facilitates integrating a variety of smart sensors to monitor patient condition, medication levels, and even recovery speed to optimize care recommendations.

To learn more, visit the Informatica Enterprise Streaming product page.