

Informatica Edge Data Streaming

Benefits

- Enable real-time operational intelligence and big data analytics
- Deliver reliable high performance streaming data collection over LAN/WAN
- Provide highly available quality of service with guaranteed message delivery
- Adapt quickly to new streaming data between multiple sources and targets
- Simplify configuration, deployment, administration and monitoring
- Increase operational efficiency and lower costs

Streaming data collection requires a new architecture.

Businesses today have an unprecedented opportunity to gain insight from a steady stream of real-time data—for example, clickstreams from web servers, application and infrastructure log data, real-time systems, and data coming from sensors or agents placed on the almost endless variety of devices and machines comprising the Internet of Things. This almost continuous flow of small messages and events can drive decision making and operational intelligence to new heights of agility and responsiveness.

However, as many small pieces of data flow in at high rates and accumulate quickly into large volumes, organizations can only derive maximum value from it if they can gather and analyze it immediately. Traditional file-based, batch-oriented collection architectures are not well-suited for streaming data because they use intermediate collection tiers, don't support real-time transmission, need to be carefully managed to prevent processing jams, and do not scale easily. They also require additional infrastructure spend and introduce inefficiency by creating multiple copies of data.

Open-source software alternatives for streaming data collection are equally inadequate, and not just because they often require high levels of software development expertise. Open-source solutions are limited in their ability to scale or adapt to enterprise-grade volumes, speeds, and variances of real-time data. Moreover, many open-source alternatives use brokers, require manual configuration, primarily collect log data, lack management and monitoring functionality, or otherwise fail to meet key requirements for a truly purpose-built streaming data collection architecture.

A new architecture for streaming data collection

Informatica Edge Data Streaming efficiently collects all forms of streaming data and delivers it directly to both real-time and batch processing technologies so companies can leverage it for holistic operational intelligence and big data analytics.

Edge Data Streaming is a distributed, scalable system that uses Informatica's established, high-performance brokerless messaging technology to greatly simplify streaming data collection through:

- Lightweight agents for an ecosystem of sources and targets
- Brokerless messaging transport using a publish/subscribe model
- Flexibility to connect sources and targets in numerous patterns
- High performance delivery direct to targets over LAN/WAN
- Simplified configuration, deployment, administration and monitoring

The provided source and target agents perform streaming data collection and distribution through the high performance message bus. The embeddable agents on sources collect data in real time and stream millions of records per second into big data platform targets such as Kafka, Hadoop, and Cassandra. The Kafka connector can be used to plug in streaming technologies such as Spark Streaming. Edge Data Streaming also streams data directly into Informatica PowerCenter Real Time Edition, Informatica RulePoint (CEP), and Amazon Web Services, enabling real-time event processing and operational intelligence.

Key Features

High performance streaming data collection with reliable qualities of service

Built on the industry's fastest and highest-performing brokerless messaging technology, Edge Data Streaming uses a publish/subscribe model that delivers data directly from source to target based on message topics of interest, with no intermediate data staging.

It leverages the underlying network to perform functions normally assigned to a broker, and can collect and deliver streaming data locally or globally over LAN and WAN. Edge Data Streaming streams data with reliable quality of service that ensures delivery as long as both source and target are up and running. For guaranteed level of service, Edge Data Streaming includes the option of adding persistent data stores in parallel to the data stream. These save data and deliver it to failed targets when they return online, with no degradation of performance.

Wide variety of supported sources and targets

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 (see Fig. 3). 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.

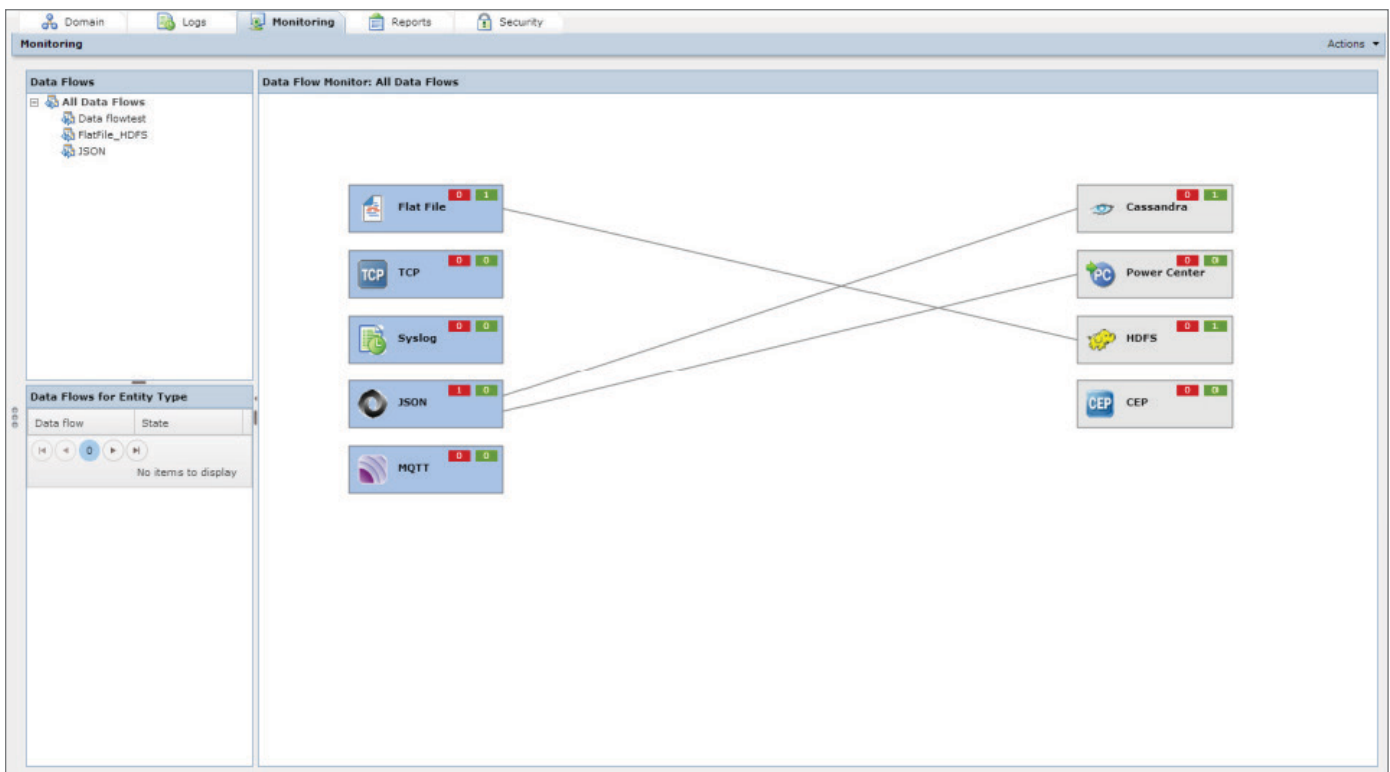
Informatica also provides an open SDK for customers who want to develop their own agents.

High availability, scalability and architectural flexibility

Edge Data Streaming is designed for high availability, with brokerless messaging architecture for fewer potential points of failure, automated failover configuration on commodity hardware with no need for a shared file system, and guaranteed data delivery using the solution's parallel persistence capability. Increasing horizontal and vertical scalability is as easy as deploying more data collection or data delivery agents and assigning them to appropriate data topics. And because sources and targets can be connected in any pattern that suits business and analytical needs—one-to-one, one-to-many, many-to-one, and many-to-many—the architecture can flex easily with changing business requirements.

Centralized GUI for simplified configuration, deployment, administration and monitoring

Edge Data Streaming uses the Informatica Administrator Console as a centralized interface for simplified configuration, deployment, administration and monitoring. A single console allows users to create flexible configurations and visual mappings for multiple source-to-target patterns; auto-generate configurations for messaging, source agents, and target agents; automate deployment; and load balance message delivery across targets to minimize impact on performance. It also provides a central point for monitoring ongoing operations and performing administrative functions like controlling agents or managing logs. This single interface facilitates ease of use, provides a complete view of the Edge Data Streaming environment, and makes it easier to add new streaming data sources or processing targets quickly.



A centralized GUI simplifies setup, deployment and monitoring, allowing users to specify the entire system topology configuration and auto-deploy agents to begin operations.

About Informatica

Digital transformation is changing our world. As the leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead the way. To provide you with the foresight to become more agile, realize new growth opportunities or even invent new things. We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption. Not just once, but again and again.

Key Benefits

Enable real-time operational intelligence and big data analytics

By collecting and delivering streaming data to big data, operational intelligence, and enterprise data integration systems, companies can achieve a truly holistic real-time picture of business activities. Edge Data Streaming helps companies transform data to intelligence faster so they 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.

Deliver reliable high performance streaming between multiple sources and targets

Vibe Data Stream streams data with highly available reliable quality of service and guaranteed message delivery. The broad array of agents built into Edge Data Streaming supports your organization's ability to monitor operations in real time through multiple sources of streaming data, even as new sources and targets emerge. The agents are easy to deploy, new ones are constantly under development, and the optional open SDK allows your IT team to develop custom agents—all ensuring that your investment in streaming data is easily expandable and future-proofed.

Simplify development and increase operational efficiency

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. Because it simplifies the direct data flow between sources and targets for operational intelligence, big data analytics and enterprise data warehousing, 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.

