

Informatica Big Data Streaming

Benefits

- Enable real-time operational intelligence with big data 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 Big Data Streaming allows organizations to prepare and process streams of data and uncover insights while acting in time to suit business needs. It can scale out horizontally and vertically to handle petabytes of data while honoring business service level agreements (SLAs).

Big Data 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.

Multilatency Data Flows Built to Last as Technologies Evolve

Informatica Big Data Streaming builds on the best of open source technologies in an easy-to-use, enterprise-grade offering. It primarily uses Spark Streaming, one of today's more vibrant open source technologies, under the covers for stream processing and supports other open source projects such as Kafka and Hadoop. As new technologies inevitably evolve, Informatica Big Data 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.

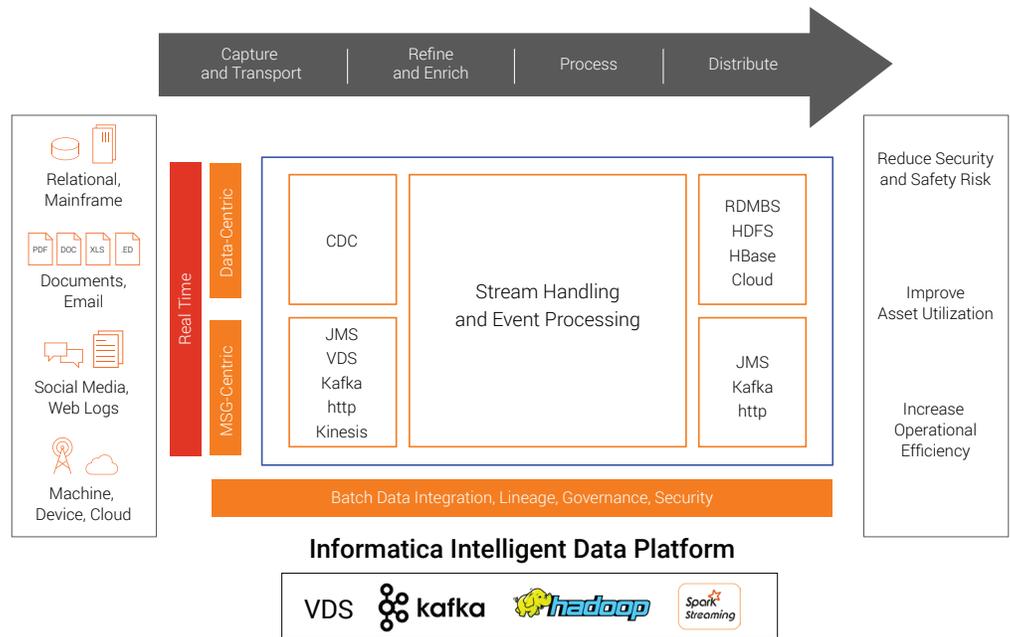


Figure 1. The Informatica Big Data Streaming flexible architecture is built on top of Informatica Intelligent Data Platform and the best of open source technologies.

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 Hadoop 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, it 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 Multitenancy 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 Big Data 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).

Cloud-Ready Streaming

More and more data is moving to the cloud to store, process, and manage real-time streaming data. Big Data Streaming 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.

Simple, Centralized Configuration, Administration, and Monitoring

Informatica Big Data Streaming is built on 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 Big Data 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.

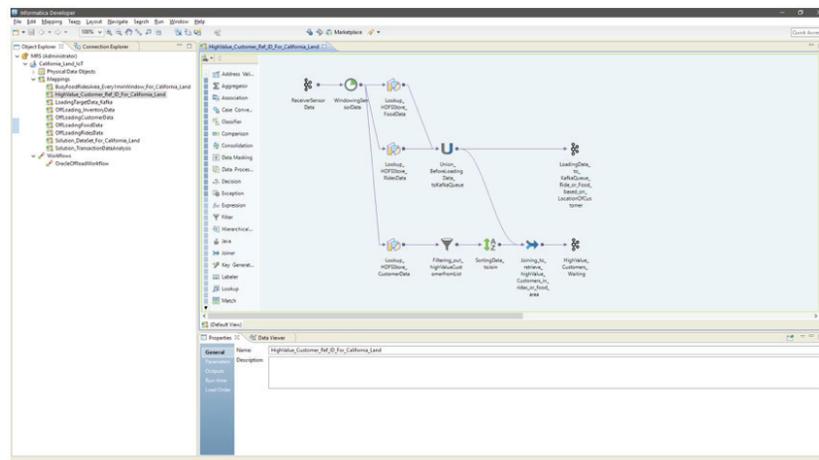


Figure 2. The Informatica Big Data Streaming visual development environment provides up to five times the productivity of hand coding.

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.

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, Multilatenency 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 Big Data 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 Big Data 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.

To learn more, see the [Informatica Big Data Streaming product page](#).



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

IN06_0418_03236

© Copyright Informatica LLC 2018. Informatica, the Informatica logo, and Informatica Intelligent Data Platform are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners. The information in this documentation is subject to change without notice and provided "AS IS" without warranty of any kind, express or implied.