



Highlights

- Perform advanced real-time analytics on data in motion
 - Rapidly ingest, correlate and continuously analyze a massive volume and variety of structured and unstructured streaming data as it arrives from thousands of sources
 - Make real-time predictions and discoveries as data arrives
 - Visualize data easily with drag-and-drop development tools
 - Detect and respond to critical events immediately
 - Learn and update models for future analysis and trend prediction with cognitive computing
-

IBM InfoSphere Streams

Real-time analytic processing for actionable insight

Every day, consumers and businesses generate data at a pace that would have seemed unprecedented just a few years ago. For example, in an average day, AT&T transfers about 30 petabytes of data through its network.¹ Cars can produce 1.3 gigabytes of sensor data per hour and, with 60 million cars manufactured per year, it is estimated that 103 exabytes of data are created just from cars.² Twitter processes 340 million tweets a day, most of which are from mobile users.³ By 2016, annual Internet traffic is expected to reach 1.3 zettabytes,⁴ and 80 percent of that data will be in unstructured formats.

Faced with this growing volume of constantly changing data, organizations are challenged to make informed, real-time business decisions and stay ahead of the competition. However, savvy organizations are fast realizing that they can extend the value of their existing systems and generate significant business advantages by updating their data mining methods to support untraditional, unstructured data sources such as audio, video and email. This enables them to react more quickly to changes in customer sentiment, uncover new market opportunities and introduce groundbreaking new products aligned with the latest trends.

As a key enabler for this new generation of analytic processing methods, IBM® InfoSphere® Streams provides a state-of-the-art computing platform that can help organizations transform burgeoning data into actionable information and business insights. InfoSphere Streams is a critical component of the IBM big data platform and delivers a highly scalable, agile software infrastructure to perform in-motion analytics on a wide variety of relational and non-relational data types at unprecedented volumes and speeds—and from thousands of real-time sources. With InfoSphere Streams, organizations can capture and act on key business data just in time, all the time.



A new paradigm for information processing

InfoSphere Streams is the result of pioneering work from IBM Research done in conjunction with the United States government. The software provides a development platform and runtime environment that enable organizations to develop and execute applications that ingest, filter, analyze and correlate massive volumes of data from continuous data streams. These data streams can originate from structured or unstructured data sources and may include a huge variety of digital information, such as:

- Text files, spreadsheets, images, video and audio recordings
- Email, chat and instant messages, web traffic, blogs and social networking sites
- Financial transactions, customer service records, telephone usage records and log files
- Data from satellites, GPS tracking, smart devices and network traffic sensors

InfoSphere Streams fuses these heterogeneous data types onto a computing platform that supports powerful data analysis with exceptional performance and impressive response times.

Powerful, real-time analytic processing made simple

InfoSphere Streams addresses a major data challenge: analyzing large volumes of data in motion. The ability to analyze this continuously generated data is often critical for organizations that must react in real time to market alerts or events, or when organizations must filter huge amounts of data and identify the rich, high-value information before moving it into a data warehouse or Apache Hadoop system.

To cost-effectively analyze streaming data, InfoSphere Streams uses cutting-edge technology like its unique Streams Processing Language (SPL). InfoSphere Streams applications can be extended with C, C++ or Java applications to reuse existing logic and speed time-to-value. In addition, the Getting Started feature makes it easy to install, build, configure and manage application instances with just a few clicks. Drag-and-drop visual development helps reduce the learning curve and speed application development.

Key components of InfoSphere Streams

Comprehensive development tools



- Eclipse IDE
- Web console
- Drag-and-drop editor
- Instance graph
- InfoSphere Streams visualization
- InfoSphere Streams Debugger

Scale-out architecture



- Clustered runtime for near-limitless capacity
- Red Hat Enterprise Linux, CentOS, SUSE Linux Enterprise Server for x86 and POWER hardware
- Support for InfiniBand and Ethernet

Sophisticated analytics with toolkits and accelerator



- CEP, database, finance, IBM InfoSphere Data Explorer, IBM InfoSphere DataStage®, geospatial, internet, message queues including MQ, MQTT and JMS. mining, R, IBM SPSS®, text, time series toolkits
- Application accelerators: telecommunications event data analytics, social data analytics

Enterprise integration: Extending the power of InfoSphere Streams

Out-of-the-box integration with other products in the IBM big data platform and built-in XML support give InfoSphere Streams access to a wide variety of data sources and systems.

- **IBM InfoSphere BigInsights™** lets you store streaming data in an enterprise-class Hadoop environment for additional analysis or historic retention. InfoSphere Streams and InfoSphere BigInsights use the same advanced text analytics capabilities to simplify natural language processing applications for both data in motion and data at rest. In addition, InfoSphere BigInsights can be used to augment streaming sources with contextual information, and users can visualize InfoSphere Streams data in the InfoSphere BigInsights console.
- **IBM InfoSphere Data Explorer** enables users to visualize InfoSphere Streams data in the InfoSphere Data Explorer CXO dashboard and add streaming data to the InfoSphere Data Explorer index.
- **IBM InfoSphere DataStage** helps users perform deep analysis and gain additional insight using contextual and source data from other parts of the infrastructure.
- **Support for XML** allows developers to fuse a broader range of traditional and untraditional data.
- **Database and warehouse support** lets users fuse and analyze streaming and unstructured data together with structured sources including IBM DB2, IBM IDS, IBM SolidDB, Microsoft SQL Server, Oracle Database, HP Vertica, IBM PureData Systems, Teradata and more.
- **Message queues with IBM WebSphere MQ, IBM MessageSight and IBM Integration Bus** provide a flexible system for efficient transport of messages and data.

Scale-out architecture

InfoSphere Streams helps organizations extend their current IT investments without a massive infrastructure overhaul. It scales from a single node to a virtually unlimited number of nodes to process data of any volume—from terabytes to zettabytes. InfoSphere Streams provides a clustered runtime environment that can easily handle up to millions of events per second with microsecond latency. Actionable results can be achieved with near-zero latency.

The advanced compiler fuses parts of the application for increased performance. The fused parts can then be distributed to run across multiple hardware nodes. The compiler also supports multiple high-speed transports, including Ethernet and InfiniBand. Existing applications can be dynamically extended with new applications to read the same streams of data, further leveraging current investments.

A web-based management console makes it easy to configure and manage the runtime and applications, including automatically placing features and deploying application components. Applications and their individual elements can be monitored for status and performance metrics to help ensure the organization attains its service-level agreements.

InfoSphere Streams includes several features that enhance high availability and redundancy. Administrators can add or remove processor nodes to or from the cluster. The ability to seamlessly add or remove nodes to or from the cluster allows administrators to perform necessary maintenance without shutting down the InfoSphere Streams application. This ability can help improve overall availability of the environment. Also, when the management nodes that control the system fail, the application nodes continue to run the InfoSphere Streams applications. The management nodes can be restarted by the administrators later.

Programmatic manipulation of streaming data makes it easy to create highly available applications by using redundant application components. Dynamic allocation of host pools and more location and isolation constraints of nodes in the runtime cluster can be used to isolate redundant application components to separate nodes in the run time.

Comprehensive tools for an agile development environment

With InfoSphere Streams, developers can easily build applications to analyze high-volume, high-velocity and high-variety data in real time.

InfoSphere Streams Studio, an Eclipse-based integrated development environment (IDE), supports rapid development with editors, wizards, application flow graphs and runtime monitoring to simplify the process of building and managing InfoSphere Streams applications. These tools are designed to help organizations quickly and easily extend their applications' capabilities with sophisticated real-time data analytics:

- **InfoSphere Streams Debugger** provides an interactive debugger for debugging SPL applications that focuses on examining the flow of data in and out of SPL operators.
- **The drag-and-drop graphical editor** allows users to build applications while automatically synchronizing graphical and SPL source code views. This round-tripping process lets developers create an application with a text or graphical editor, and update later using the other editor.
- **An instance graph**, available in both InfoSphere Streams Studio and the management console, provides visual monitoring of application health and metrics and lets users quickly identify issues using customizable views.
- **The latest InfoSphere Streams data visualization capabilities** allow users to dynamically add new views to running applications with charts provided out of the box.

InfoSphere Streams: System requirements

Hardware requirements	<ul style="list-style-type: none"> • Intel/AMD x86 architecture (64-bit) or IBM POWER7® architecture systems (minimum 500 MB memory) • 2 GB memory to run simple applications
Software requirements	<ul style="list-style-type: none"> • Red Hat Enterprise Linux (RHEL) Version 5.6 or later, or Version 6.1 or later for x86 architecture hardware • RHEL Version 6.2 or above for IBM POWER7 architecture hardware • CentOS Version V5.6 or later, or 6.1 or later for x86 architecture hardware • SUSE Linux Enterprise Server (SLES) V11.2 or above for x86 architecture hardware (SLES does not support Security Enhanced Linux) • Eclipse Platform SDK V4.2.2 • IBM Java SE V6 or V7 SDK for x86 or IBM POWER7 architecture hardware • Oracle Java SE v6.0.x or V7.0 SDK for x86 architecture hardware • Mozilla Firefox v10 or above • Microsoft Internet Explorer 9.0 or above

Sophisticated analytics with toolkits and accelerators

InfoSphere Streams features integrated toolkits and sample applications that facilitate the development of solutions for particular industries or functions. These include the data mining toolkit, financial services toolkit and standard toolkit, which encompass an array of the most commonly used operators for smooth integration with existing infrastructures and applications.

For example, a financial services organization can use InfoSphere Streams to combine incoming stock market quotes with industry data or information on interested traders that is held on an internal database. Adding context to stock market ticks helps the organization achieve more complex levels of analysis and customized alerting than ever before. In addition, InfoSphere Streams helps determine, in real time, which data should be stored in a database for future analysis, reducing storage and administration costs.

Other toolkits in InfoSphere Streams include:

- **Complex Event Processing (CEP)**, which offers rich, high-performance analytics by using patterns to detect composite events in streams of simple events. Existing applications are easily integrated and can be migrated into an InfoSphere Streams environment to take advantage of increased scalability and the ability to process up to 10 times more events per second on the same hardware.
- **InfoSphere Data Explorer**, which brings visualization and storage capabilities to InfoSphere Streams by enabling users to pass visualization widgets from InfoSphere Streams to InfoSphere Data Explorer, or pass data to InfoSphere Data Explorer for indexing and later retrieval.
- **InfoSphere DataStage** allows users to perform in-depth data analyses as part of the information integration flow and see rapid results. An InfoSphere Streams toolkit provides adapters that exchange data between InfoSphere Streams and InfoSphere DataStage for real-time analytics processing.
- **IBM SPSS** leverages IBM SPSS Modeler to develop and build predictive models, and then deploy them using the SPSS Scoring Operator. SPSS models can be dynamically refreshed in InfoSphere Streams without suspending InfoSphere Streams applications.
- **Geospatial** is often used in the transportation and telecommunications industries. It enables location-based marketing based on high-performance analysis and processing of geospatial data.
- **Messaging queues** allow InfoSphere Streams to receive data from or send data to IBM WebSphere® MQ, IBM MessageSight and Java Messaging System (JMS) offerings.
- **Time series** is a rich set of tools that includes data generation (synthesizing or extracting), preprocessing (preparation and conditioning), analysis (statistics, correlations, decomposition and transformation), and cognitive computing and modeling (prediction, regression and tracking).

InfoSphere Streams also includes application accelerators that help improve time-to-value and leverage best practices around a specific solution implementation or use case. They provide business logic, data processing and UI/visualization capabilities, analytic operators or sample applications that can be used as design templates for new applications.

The IBM Accelerator for Social Data analytics provides rapid implementation to analyze large volumes of various types of social media data with real-time processing. This helps organizations gain vital insight for effective, highly targeted marketing campaigns and timely product and marketing decisions. It also facilitates insights that support competitive intelligence efforts and building customer retention and new-customer acquisition programs.

The IBM Accelerator for Telecommunications Event Data Analytics provides transformation and analytics for telephone companies' call and event-detail records, which can help with real-time revenue assurance and fraud detection. This helps telecommunications providers strengthen billing insights based on services, vendors and business lines. It also supports efforts to differentiate services, strengthen customer loyalty, reduce churn, offer targeted services, personalize billing and improve the customer experience.

InfoSphere Streams V3.2: Features and enhancements

The latest version of InfoSphere Streams includes a wide range of new features and functionality enhancements to improve usability, support enterprise readiness and streamline enterprise integration.

Improved usability

InfoSphere Streams is simple to use because it comes with an impressive suite of development tools, sophisticated analytics toolkits and accelerators all designed to make it easy to develop applications and work with a wide variety of data in motion. InfoSphere Streams v3.2 expands the development platform by providing remote development for Windows and Linux, user-defined parallelism and Java simplification. No other platform is easier to install, develop, maintain, monitor and scale. You can now use Streams Studio natively on a Windows desktop to develop applications. Streams Studio interacts seamlessly with Linux-based cluster instances to test and deploy applications.

InfoSphere Streams v3.2 delivers user-defined parallelism. Developers can replicate any part of an application any number of times to perform parallel processing, and split the streaming records using either round-robin or hash-based algorithms. The amount of parallelism can be specified at compile time. Developers can also force parallel channels to run on different hosts to improve application performance. Finally, Java operator model simplification means developers can maintain implementation source code and the model together more easily.

Enhanced enterprise readiness

Already enterprise-ready, InfoSphere Streams delivers a scale-out architecture and clustered runtime for near-limitless capacity. InfoSphere Streams v3.2 makes it even simpler to stay up and running while dynamically adding rules and interfaces.

InfoSphere Streams v3.2 enables users to access application data programmatically using Representational State Transfer (REST) interfaces from any portion of the application. Support for the IBM Operational Decision Management (ODM) delivered in InfoSphere Streams v3.2 enables execution of ODM rules as part of an application. Users can create, add, import, and update ODM rules at runtime without having to restart applications.

Streamlined enterprise integration

InfoSphere Streams is designed to be compatible with a variety of software and tools. In InfoSphere Streams 3.2, new adapters and operators further extend InfoSphere Streams integration and make integration quicker and easier. For example, the Message Queueing Telemetry Transport (MQTT) adaptor is now supported. MQTT is a machine-to-machine (M2M) and Internet of Things (IoT) connectivity protocol designed as an extremely lightweight publish and subscribe messaging transport. It is useful for connections with remote locations where a small footprint is required or network bandwidth is at a premium, such as mobile devices or automotive telematics applications. InfoSphere Streams v3.2 also adds BigInsights BigSQL and HP Vertica support.

Changing the face of big data analytics

With InfoSphere Streams, organizations can see events and trends as they are happening and respond proactively. By bringing meaning to unwieldy, vast and fast-moving data streams, the software has already delivered real benefits to organizations in a wide range of industries.

For example, a communications company is currently using InfoSphere Streams to mediate more than 200,000 call-detail records per second and simultaneously create summaries of that data. The summary process formerly took at least 12 hours, but now, with summaries available almost immediately, the company can make critical decisions based on real-time call information. This allows them to target high-value customers with personalized marketing offers based on their current location, when it is still possible to influence the customer.

In the financial services sector, an InfoSphere Streams-based application analyzes and correlates over five million market messages per second to execute algorithmic option trades with an average latency of 30 microseconds, giving the business the agility to take advantage of market trends more quickly than the competition. And in healthcare, InfoSphere Streams is being used to analyze information from multiple sensors in a neonatal intensive care unit, with the goal of being able to detect changes in patients' medical conditions up to 24 hours earlier than before—a decided edge in treating illness successfully.

For more information

Any organization that requires immediate, accurate analysis and business decisions based on up-to-the-minute information can benefit from InfoSphere Streams. To learn more about InfoSphere Streams and how it can help your business gain a competitive advantage, visit:

ibm.com/software/data/infosphere/streams

Get it now! Real-time analytic processing at your fingertips: Introducing IBM InfoSphere Streams Quick Start Edition

Free to download, quick to start. To help clients get started with real-time analytic processing of data in motion, IBM has launched the Quick Start program, which gives users fast access to market-leading big data technology. Available through a free download, the market leading, enterprise-grade big data platform empowers users to experiment with stream computing in their unique, non-production environment. There is no data or time limit. [Read more on the InfoSphere Streams Quick Start Edition website.](#)



© Copyright IBM Corporation 2013

IBM Corporation

Software Group
Route 100
Somers, NY 10589

Printed in the United States of America
October 2013

IBM, the IBM logo, ibm.com, BigInsights, DataStage, InfoSphere, Power, POWER7, SPSS, and WebSphere are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml

Intel, Intel logo, Intel Inside, Intel Inside logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

1 Kahn, Mobeen. “The Big Data Boom: How Mobile and Fixed Devices Are Feeding Big Data and What It Means to Businesses,” AT&T Networking Exchange blog, July 19, 2012. <http://networkingexchangeblog.att.com/enterprise-business/the-big-data-boom/>

2 <https://twitter.com/MobileBit/status/385162303580016640Dublin>

3 “Twitter turns six,” Twitter blog, March 21, 2012. <http://blog.twitter.com/2012/03/twitter-turns-six.html>

4 Moscaritolo, Angela. “Internet Traffic to Reach 1.3 Zettabytes by 2016,” PC Magazine, May 30, 2012. <http://www.pcmag.com/article2/0,2817,2405038,00.asp>

5 Runtime performance increase for operations using bounded lists and maps compared to InfoSphere Streams 3.0.



Please Recycle