



Highlights

- Acquire, analyze and act on a broad range of data in motion and make decisions while events occur
 - Automate deployment of stream processing applications to cloud or on-premises environments
 - Unlock hidden insights from human voice by transforming speech to text in real time
 - Build a federated Internet of Things (IoT) application leveraging analytics at the edge with Apache Edgent
 - Develop using a variety of languages, including Streams Processing Language (SPL), Java, Scala and Python
-

Real-time analytics with IBM Streams

The streaming analytics engine of the cognitive business

Organizations need data and analytics to remain competitive

Organizations are often overwhelmed by the volume and variety of data moving through their organizations at high speed. Simply collecting or managing that big data shouldn't be the end goal. New sources of data—such as the IoT and rapidly changing sentiments aggregated from social sites—can't easily be captured in a traditional data center.

To innovate, organizations must quickly make sense of all this data—from persistent structured data to streaming data from sources such as smart appliances or wearable devices in the IoT—decisively, consistently and in real time. Understanding the *context* of data is key to improving customer relationships, enhancing operational efficiencies, reducing risk and uncovering new opportunities.

However, as the collective ability to observe grows, making sense of those observations becomes exponentially more difficult. The result is often a waste of capital and operational resources. Too much time and effort is spent chasing false opportunities, while real ones go unrealized.

According to the seventh annual IBM Institute for Business Value analytics global study, a majority of organizations are embracing these marketplace disruptions as opportunities, intensifying their adoption of big data and analytics technologies.¹ According to 95 percent of the 1,226 respondents surveyed, big data and analytics capabilities are now necessary to keep up with competitors or outperform them.



Competitive impact of data and analytics on organizations	Percentage of responses from global organizations	Percentage of responses from local organizations	Total percentage
Keeps them ahead of competitors	14%	33%	47%
Keeps pace with competitors	25%	23%	48%
No advantage versus competition	N/A	N/A	5%

Table 1: Data and analytics have a global or local impact on 95 percent of all organizations.

Source: IBM Institute for Business Value 2015 Analytics research study, "Analytics: The upside of disruption"

Streaming data is a challenge

Every day, consumers and businesses generate data at unprecedented speeds. For example, the ICU at Emory University Hospital generates 100,000 streaming data points per second, running up to 50 algorithms on each patient continuously.² Telecommunications providers are analyzing 7 billion call data records (CDRs) per day.³ Social media also contributes to this massive data growth. Every second, on average, 6,000 tweets are sent using Twitter, which corresponds to more than 350,000 tweets per minute, 500 million tweets per day and around 200 billion tweets per year.

Organizations need to identify risks and opportunities in high-velocity data—opportunities that often can be detected and acted on only at a moment's notice. Flows of data from streaming sources, such as market data, the IoT, mobile devices, sensors, clickstreams and even transactions, remain largely un navigated. It's time to unlock this data to optimize decision making.

IBM Streams acts on all your data in real time

IBM® Streams analyzes the broadest range of streaming data, making decisions while events are happening. Streams brings meaning to fast-moving data streams and helps organizations in a wide variety of industries. Now organizations can subscribe to device data to provide advanced analytics using Streams with IoT platform capabilities.

As part of a new generation of analytic processing methods, Streams offers a computing platform that helps organizations turn burgeoning, fast-moving volumes and varieties of data into insight. It delivers a programming language and an integrated development environment (IDE) for applications, a runtime system that executes the applications on a single host or a distributed set of hosts and analytic toolkits to help speed development. You can use Java and Python to develop applications for deployment to the runtime. Streams can ingest, filter, analyze and correlate massive volumes of continuous data streams.

These data streams can originate from any of the following:

- IoT devices and sensors
- Text files, spreadsheets, images, video and audio recordings
- Email, chat and instant messaging; web traffic, blogs and social networking sites
- Financial transactions, customer service records, telephone usage records, and system and application logs
- Satellite data, GPS data, smart devices, sensors, network traffic and messages

The fundamentals of IBM Streams

- Acquire the broadest range of data types
 - Apply continuous, multimodal analytics
 - Act at the appropriate time, using the most appropriate method
-

The capabilities of IBM Streams

Streams is designed to deliver breakthrough capabilities that enable aggressive analysis and management of information and knowledge. Streams is highly efficient, using 14.2 times fewer hardware resources and delivering 12.3 times more throughput compared to open source offerings.⁴

A key component of the IBM Watson® Data Platform, Streams processes huge volumes and varieties of data from diverse sources with extremely low latency. This component enables decision makers to extract relevant information for timely analysis.

Streams radically extends the state of the art in information processing by helping organizations to:

- Continuously analyze data in motion across multiple sources to deliver actionable insight.
- Connect to any data stream to make predictions and discoveries as data arrives to enhance and improve analytic models and cognitive systems.
- Deploy a complete set of streaming analytics—such as natural language processing, geospatial, predictive and more—to satisfy unique, industry-specific requirements and use cases.
- Speed time to value with open-source technologies through the use of Java APIs and visualize data easily with drag-and-drop development tools that help support faster time to deployment and effective production management.

- Detect and respond to critical events 10 times faster compared to traditional techniques to stay ahead of the competition, analyzing millions of interactions or events per second.
- Administer with both web-based and command-line interfaces for optimal enterprise management strategies.
- Integrate with Spark and Hadoop to capture data streams efficiently and use them alongside data at rest within the organization.
- Increase confidence in data stream insights through integrated data lineage and governance.

Choose cloud or on-premises deployment to help manage costs and further speed deployment. While certain research, open source and commercial initiatives try to address these technical challenges individually, Streams is designed to simultaneously address all of them. This integrated approach enables Streams to break through a number of fundamental barriers. In fact, Forrester named IBM a leader in The Forrester Wave™ for Big Data Analytics Platforms.⁵

A paradigm shift from queries to continuous analytics

Stream processing extends traditional approaches to information processing, such as transactional or complex event processing (CEP) systems. Traditional processing involves running analytic queries against historic data.

Let's use an app that calculates the distance users have walked each month using global positioning system (GPS) location data as an example. The traditional approach would make one calculation using all the data transmitted from the user's mobile device and stored over the past month. In contrast, stream processing performs continuous analytics to keep running totals that are updated moment by moment as the GPS data is refreshed.

In the first case, questions are answered using historic data; in the second case, answers are continuously updated using streaming data. See Figure 1. Stream processing differs from a simple in-memory database in that traditional analytics systems must first load all the data—even if in memory—and then run a query.

Streams augments the traditional analytics approach by enabling continuous analysis to be modified over time. For example, it can use machine-learning techniques to determine the most popular walking routes or detect deviations from a user’s standard walking patterns.

To support this fundamentally different approach to continuous analysis, Streams offers a distributed runtime platform, a programming model and tools for developing continuous analytic applications.

The data streams consumed by Streams can originate from satellites, sensors, cameras, news feeds, stock tickers and a variety of other sources, including traditional databases and Apache Hadoop systems.

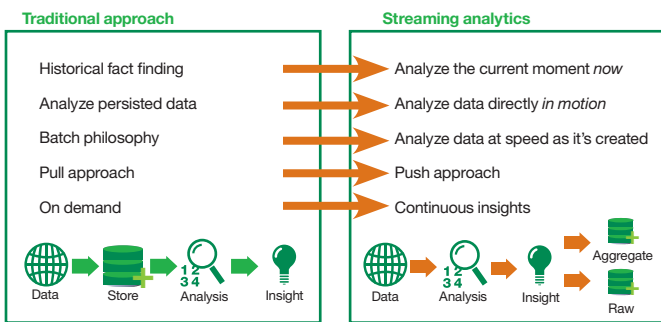


Figure 1: Comparing the traditional approach to analytics with streaming analytics.

What's new in IBM Streams v4.2

Streams v4.2 is poised for continued growth with new Python and Rules developer support, IoT edge analytics, and Hyperstate Accelerator. Streams v4.2 offers the following benefits:

- Python development: Developers can create streaming applications in Python.
- Speech-to-text toolkit: This toolkit enables Streams to analyze English language speech and convert it to text for further analysis and insight.
- Streams Rules: This rules compiler enables original design manufacturing Rules to run natively on Streams and is designed to provide superior performance and low latency.
- Edge-to-Center Analytics: Support for Apache Edgent, formerly Quarks. Java-based streaming analytics, targeted at the IoT market, delivers analytics to the system’s edge.
- Hyperstate Accelerator: With its high-speed Hyperstate Accelerator, Streams natively supports state management and event playback, which is designed to process at extremely high speed.

Faster streaming application delivery

Time to value is crucial to businesses and one of the methods to speed time to value is to reduce the learning curve when new solutions are implemented. Streams provides faster streaming application delivery through the use of Python. Python is among the preferred programming languages for writing algorithms and machine learning.

Many data scientists and engineers that develop analytics applications are likely to know Python, helping to reduce the learning curve. With that reduced learning curve also comes faster application delivery and a faster time to value for the business overall. Developers can use existing Python code for building Streams applications without starting from scratch.

Unlock hidden insights from the human voice

Streams, using IBM Watson speech to text, converts spoken English to text. With Streams v4.2, developers can now transform speech to text in near real time, which can then be analyzed by Streams for further insights. The speech-to-text toolkit enables developers to create applications that ingest voice, convert it to text and then, with the Text Toolkit, perform natural language processing.

The speech-to-text service can be used in nearly any scenario where voice interactivity is needed. This feature is designed for mobile experiences, transcribing media files, call center transcriptions, voice control of embedded systems or converting sound to text, enabling data to be searchable.

Speech analytics is the key to unlocking hidden insights by analyzing the wealth of information from customer calls for improved customer satisfaction.

Build a federated IoT application leveraging analytics at the edge with Apache Edgent

Because the IoT is increasingly important to the business world, the Streams v4.2 update also integrates Apache Edgent for IoT analytics at the edge.

New support for Apache Edgent allows Streams developers to create federated applications to optimize computing for IoT applications with Edgent at the edge, and Streams for central analytics. Developers performing analytics on the edge can use the Streams integration with Edgent to help manage and control analytics through the Streams console.

Streams provides functions that help simplify connecting IoT applications and submitting them to Streams using the Streams console. The Streams application integrates with the IBM Watson IoT™ platform to allow developers to receive device events and status updates, and send device commands without having to understand MQ telemetry transport (MQTT) topics and connectivity details.

Simplified view of the application

Streams also provides a more simplified view of the application by creating rules using an industry-leading solution, IBM Operational Decision Management. With this solution, the developer creates rules using the same language as the Operational Decision Management capabilities.

Streams already has support for Operational Decision Management business rules with the rules toolkit, introduced in a previous release, and called and ran rules in the solution. The new capability compiles the rules into Streams Processing Language for native execution, so licensing of the solution is no longer required.

“Through our ongoing partnership and engagement with IBM, we now enable artificial intelligence and machine learning, predicting customer intent by mining unstructured data and correlations.”

— Damu Bashyam, Vice President of IT, Verizon

Cognitive customer experience platform

Verizon worked closely with IBM to create the Verizon cognitive customer experience platform. This platform includes continuous ingestion of several data sources, high-performance analytical microservices and decision engines, with near real-time scoring of customer interactions. The software stack includes IBM Watson speech to text, IBM Watson Explorer, as well as the IBM Watson Data Platform, which uses IBM Streams and open source components.

Watch the client voices channel at YouTube.com:

ibm.co/2gCIR2a

Read the Verizon case study at:

ibm.com/software/businesscasestudies/us/en/corp?synkey=T684679Y42405E43

Enterprise integration: Extending the power of Streams

Out-of-the-box integration with other offerings gives organizations access to a wealth of data and systems.

- IBM BigInsights®: Stores streaming data in an enterprise-class Hadoop environment for additional analysis or historical retention. Users can integrate data into BigInsights to augment streaming sources with contextual information and visualize Streams data in the BigInsights console.
- IBM InfoSphere® IBM DataStage®: Enables users to perform deeper analysis and gain more insight by including contextual and source data from other parts of the infrastructure.
- XML support: A broader range of traditional and nontraditional data can be integrated for development work.
- Microsoft Excel: Business users can analyze and visualize streaming data in Microsoft Excel worksheets.
- IBM SPSS®: Uses SPSS Modeler for developing and building predictive models. This platform deploys models using the SPSS Scoring Operator. SPSS models can be dynamically refreshed in Streams without suspending Streams applications.
- Apache Edgent, previously called Apache Quarks: Create federated analytics between Streams and devices and gateways at the edge of networks running Apache Edgent. Communications can be performed natively with Streams or using a management service, such as the IBM Watson IoT platform on IBM Bluemix®.
- IBM Watson IoT platform: Use the IoT device management capabilities of the IoT platform and subscribe to device data to provide advanced analytics using Streams.
- IBM InfoSphere Information Governance Catalog: Speed development by dragging and dropping data definitions onto InfoSphere Streams Studio to generate schema. Streams can also feed lineage data into the InfoSphere Information Governance Catalog.
- IBM Watson Explorer: Enable users to visualize Streams data in the IBM Watson Explorer CXO dashboard and add streaming data to the IBM Watson Explorer index.

Stream computing industry use cases

Organizations have developed hundreds of applications for Streams in the past few years. These applications are used in diverse industries, such as telecommunications, government, financial services, healthcare, energy and utilities, insurance and automotive, among many others. See Figure 2.



Figure 2: Streaming analytics have a wide range of benefits in virtually all types of industries.

Additional resources

Join the IBM Streams Developer Community. Streams has a thriving community of developers and architects. Join now at <https://developer.ibm.com/streamsdev>

IBM is a Leader in The Forrester Wave—Big Data Streaming Analytics, Q1 2016:

ibm.com/marketing/iwm/dre/signup?source=ibm-analytics&S_PKG=ov47736&dynform=22013&S_TACT=000000VP&S_OFF_CD=10000737

Watch the webcast: Streaming Analytics Webcast by Bloor Research and IBM at: ibm.com/marketing/iwm/iwm/web/signup.do?source=ibm-analytics&S_PKG=ov53840&S_TACT=000000VP&S_OFF_CD=10000737

IBM is a Champion vendor in the Bloor Research Streaming Analytics Report 2016. Read the report at: ibm.com/marketing/iwm/iwm/web/signup.do?source=ibm-analytics&S_PKG=ov53201&S_TACT=000000VP&S_OFF_CD=10000737

Walmart and IBM Revisit the Linear Road Benchmark. View the slide presentation at: <http://www.slideshare.net/RedisLabs/walmart-ibm-revisit-the-linear-road-benchmark>

Conclusion

Organizations need to spot risks and opportunities in high-velocity data—opportunities that often can be detected and acted on only at a moment’s notice. Flows of data from streaming sources, such as market data, the IoT, mobile devices, sensors, clickstreams and even transactions, remain largely un navigated. It’s time to unlock this data to optimize decision making.

For more information

To learn more about IBM Streams, please contact your IBM representative or IBM Business Partner, or visit: ibm.com/streams or visit: ibm.com/marketplace/cloud/stream-computing/us/en-us



© Copyright IBM Corporation 2017

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
February 2017

IBM, the IBM logo, ibm.com, IBM Watson, BigInsights, Bluemix, DataStage, InfoSphere, and SPSS 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.

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

Microsoft and Excel 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.

All client examples cited or described are presented as illustrations of the manner in which some clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions. Contact IBM to see what we can do for you. It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

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.

- 1 IBM Institute for Business Value 2015 Analytics research study, “Analytics: The upside of disruption”
- 2 IBM, “Emory University Hospital uses IBM streaming analytics to gain life saving insights on patients,” 21 August 2014, ibm.com/software/businesscasestudies/us/en/corp?synkey=Y890037M62954T62
- 3 IBM, “Emory University Hospital uses IBM streaming analytics to gain life saving insights on patients,” 21 August 2014, <http://bit.ly/2joiPvo>
- 4 IBM Institute for Business Value 2015 Analytics research study, “Analytics: The upside of disruption”
- 5 The Forrester Wave - Big Data Streaming Analytics, Q1 2016



Please Recycle