



SPECIAL REPORT

STREAMING ANALYTICS: HOW TO REALIZE ITS FULL POTENTIAL

Streaming analytics is all about extracting business value from data in motion in the same way traditional analytics tools make use of data at rest.

Streaming analytics can be used to improve operational efficiencies, reduce infrastructure costs, and provide faster insights and actions.

INTRODUCTION

Organizations in every industry have data streaming available from applications, social media, sensors, devices, websites, and more. Analyzing this data in real time, rather than storing it and analyzing it later, can provide critical, actionable insights in many operational and functional areas.

Essentially, streaming analytics is all about extracting business value from data in motion in the same way traditional analytics tools make use of data at rest. Real-time streaming analytics helps companies of all sizes by issuing alerts instantly when a customer's experience is degraded, or fraud is detected.

Additionally, information derived from real-time analytics can be used to identify anomalies and business changes (a sudden spike in demand for a product or service, or a defect in manufacturing) as they occur. Such information allows companies to take instant action and seize an opportunity that they otherwise might miss.

Use cases of real-time analytics performed on streaming data can be found in cybersecurity, financial services, retail, manufacturing, the energy industry, healthcare, and many more. Some common examples of real-time analytics of streaming data include the following:

- Many manufacturers embed intelligent sensors in devices throughout their production line and supply chain. Analyzing the data from these sensors in real time allows a manufacturer to spot problems and correct them before a product leaves the production line. This improves production and efficiency of operations—and saves money.
- In cybersecurity uses, streaming analytics can instantly identify anomalous behavior and suspicious activities and flag them for immediate investigation. So, rather than remediating after a problem occurs, the attack is stopped before it can do any damage.
- In the hospitality industry, a hotel chain might use streaming analytics to monitor reservations in real time. Spotting a location that has high availability in the late afternoon, the chain could act (by texting or emailing special promotions to frequent guests, for example) to fill those empty rooms that night.

In many cases, companies use real-time streaming analytics to complement information derived from other applications, such as transaction processing. Consider a global airline that processes thousands of reservation transactions per second. Real-time analytics could optimize revenue opportunities by finding out which routes have a surplus of seats that need to be filled.



To derive actionable information from all this data, a streaming analytics solution must be able to ingest data from a variety of sources, and offer the performance and scalability needed to analyze that data in real time.

FACTORS TO CONSIDER WHEN IMPLEMENTING STREAMING ANALYTICS

Every day, consumers and businesses generate data at unprecedented speeds. The following statistics put the magnitude of this data tsunami into perspective:

- Every minute, 456,000 tweets are posted on Twitter, along with 510,000 comments and 293,000 status updates on Facebook. During that one-minute timespan, 156 million emails are sent.¹
- The Intensive Care Unit (ICU) at one leading hospital reported that it generates 100,000 streaming data points per second.²
- Taking all sources of data generation together, there are 2.5 quintillion bytes of data created each day.³

In addition, data volumes are expected to skyrocket, as the number of internet of things (IoT) devices is projected to grow to 200 billion by 2020.⁴ Much of that data will be used by companies in the course of doing business. Research firm Gartner predicts that more than half of new business processes and systems will incorporate some IoT elements by 2020.⁵

To derive actionable information from all that data, a streaming analytics solution must be able to ingest data from a variety of sources, and offer the performance and scalability needed to analyze that data in real time.

Furthermore, a streaming analytics solution should be a general-purpose solution that can be integrated with, or embedded into, other applications. Tight integration is required if the results derived are used to make real-time decisions or to take actions regarding customer service, security or fraud prevention, product recommendations, or automated chat or voice response.

One additional factor to consider is ease of use. The reason: Analytics is not used just by data scientists and business analysts anymore. All types of managers and employees—in sales, marketing, finance, and customer service—are using analytics systems and services to do their jobs more effectively. Key capabilities needed to make streaming analytics available to such decision-makers include built-in data visualization and dashboards.

EXPLORING STREAMING ANALYTICS OPTIONS

Streaming analytics can deliver real-time analysis, support event-driven digital transformation, and provide IoT analytics. Companies that want to take advantage of streaming analytics have several choices in how they deploy their solution. The options fall into these three categories:

The do-it-yourself (DIY) on-premises approach: In this situation, a company would design, evaluate, install, and manage an array of hardware and software, so it would need in-house expertise to plan, assemble, and fine-tune its streaming analytics solution. Most companies do not have this internal expertise and must either hire new staff or contract with third-party providers.

There are other issues to consider with a DIY approach. Many data centers do not have the physical space, extra electrical capacity, or cooling capabilities to support more equipment. A DIY approach incurs large CAPEX costs and recurring OPEX costs. Additionally, once a solution is implemented, the company may not be able to scale up or support new technologies.

The home-grown, open-source cloud approach: An alternative to a DIY on-premises solution is to use cloud services running open-source streaming analytics software. Today, there are many open-source solutions available, including AthenaX, Apache Flink, Apache Spark, Apache Storm, Hortonworks Streaming Analytics Manager, and others.

One point to consider with an open-source cloud approach is that companies still need in-house expertise to bring the solution together. As noted above, many businesses do not have expertise in streaming analytics since this is a relatively new field.



IBM Streaming Analytics solves the challenges of setting up and using streaming analytics by providing an easy-to-use cloud-based stream processing service.

A second consideration is that general-purpose cloud services may not offer the performance needed. Some may not be able to support the very high data ingestion rates that are common in streaming analytics applications. With others, the compute instances may not have the processing power necessary to analyze incoming data instantly.

A dedicated streaming analytics cloud approach: To meet today's demand for high-performance real-time analytics of streaming data, several dedicated cloud services have recently emerged. These services differ from general cloud services running open-source streaming analytics software in several ways. They support high data ingestion rates and are very scalable. Critically, dedicated streaming analytics cloud solutions deliver enterprise-grade availability and support.

IBM AS YOUR STREAMING ANALYTICS CLOUD PROVIDER

Selecting and deploying the right streaming analytics solution for your company is a complicated and time-consuming task. To get the most out of their deployment, many companies turn to a technology partner that offers both streaming analytics expertise and in-depth, real-world experience.

This is an area in which IBM has the expertise to provide extensive assistance. IBM Streaming Analytics solves the challenges of setting up and using streaming analytics by providing an easy-to-use cloud-based stream processing service. The solution connects to the streaming data source and target systems and is configured to perform any analytics or data processing operations needed while the data is in transit.

Specifically, companies can perform real-time analysis on data-in-motion as part of an IBM Cloud application. The Streaming Analytics service is powered by IBM Streams, which can analyze millions of events per second, enabling sub-millisecond response times and instant decision making.

IBM Streaming Analytics can handle very high data rates. With IBM Streams, a company can ingest, analyze, monitor, and correlate data as it arrives from real-time data sources. Managers and IT staff can view information and events as they unfold.

Deploying applications to a Streaming Analytics instance running in the IBM Cloud, IBM Streams can handle very high data rates and perform its analysis with predictable low latency, so that an application can operate at the speed of data.

From its initial launch in 2009, IBM Streams has been designed for enterprise deployment and includes a comprehensive set of accelerators and templates to help businesses get up and running quickly. For example, IBM can provide one accelerator to help financial markets recognize patterns in high-volume trading and another to help telecom companies process and analyze millions of call data records every second. Streams is also designed to integrate easily with an enterprise's existing data and analytics environments, such as data warehouses, data lakes, and big data platforms.

IBM Streams offers some key differentiators.

Using IBM Watson speech-to-text capabilities, Streams converts spoken words to text in near real time. This enables data to be searchable and analyzed for further insights. The speech-to-text service can be used in nearly any scenario in which voice interactivity is needed. It is ideal for mobile experiences, transcribing media files, call center transcriptions, voice control of embedded systems and converting sounds to text.

Because the IoT is increasingly important to business, Streams also integrates Apache Edgent for IoT analytics at the edge. This 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.



Additionally, IBM offers a variety of tutorials to help companies use the Streaming Analytics service with other IBM Cloud services, including messaging,⁶ database,⁷ and predictive analytics⁸ services. All the integration tutorials include code samples.

Also of note, the latest version of the IBM Streaming Analytics service on IBM Cloud has been rebuilt to run in Docker containers, instead of bare virtual machines, and it uses Kubernetes for container orchestration.

The result is greater availability because Kubernetes can adjust the environment in real time to maintain the desired service levels. For example, it can spin up new containers to keep streams online, even when a current instance of the Streaming Analytics service needs to be taken offline for patching or upgrades.

The new architecture also enables a more dynamic approach to resource allocation: Companies can specify the maximum number of nodes they want their environment to use, and the service will automatically scale up and down within that threshold. This helps ensure that a company only uses—and pays for—the resources it needs at any point in time, while still maintaining performance during peak load periods.

SUMMARY

Stream processing is becoming increasingly important for companies. In response to this growing demand, many solutions have been developed—both by traditional enterprise technology vendors and by the open-source community.

IBM Streaming Analytics, a front-runner in the marketplace, provides many benefits and differentiators, including the following:

- Provides an analysis of the broadest range of streaming data
- Allows companies to make decisions while events are happening
- Offers cost-effective pricing
- Is highly scalable
- Makes use of IBM Watson services

The bottom line is that IBM Streaming Analytics on IBM Cloud can be used to improve a company's operational efficiencies, reduce infrastructure costs, and provide faster time to both insights and actions.

For more information about IBM cloud-based streaming analytics services, visit:
<https://www.ibm.com/cloud/streaming-analytics>

¹“[How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read](#),” Forbes, May 21, 2018

²“[Healthcare predictive analytics: An overview with a focus on Saudi Arabia](#),” ScienceDirect, 2018

³“[Data Never Sleeps 5.0](#),” Domo, July 2017

⁴“[A Guide to the Internet of Things](#),” Intel

⁵“[Internet Of Things Market To Reach \\$267B By 2020](#),” Forbes, January 29, 2017

⁶“[Get Started with Streaming Analytics + Message Hub](#),” IBM, October 16, 2015

⁷“[Using Streaming Analytics with JDBC-enabled Bluemix databases](#),” IBM, January 26, 2016

⁸“[Support for SPSS Analytics Toolkit in Bluemix Streaming Analytics Service](#),” IBM, August 14, 2015



RTInsights is an independent, expert-driven web resource for senior business and IT enterprise professionals in vertical industries. We help our readers understand how they can transform their businesses to higher-value outcomes and new business models with IoT real-time analytics. We provide clarity and direction amid the often confusing array of approaches and vendor solutions. We provide our partners with a unique combination of services and deep domain expertise to improve their product marketing, lead generation, and thought leadership activity.



IBM Streams analyzes the broadest range of streaming data, making decisions while events are happening. It 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.

A key component of the IBM Cloud Platform, 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.