IDC MarketScape: Worldwide Analytic Stream Processing Software 2024 Vendor Assessment
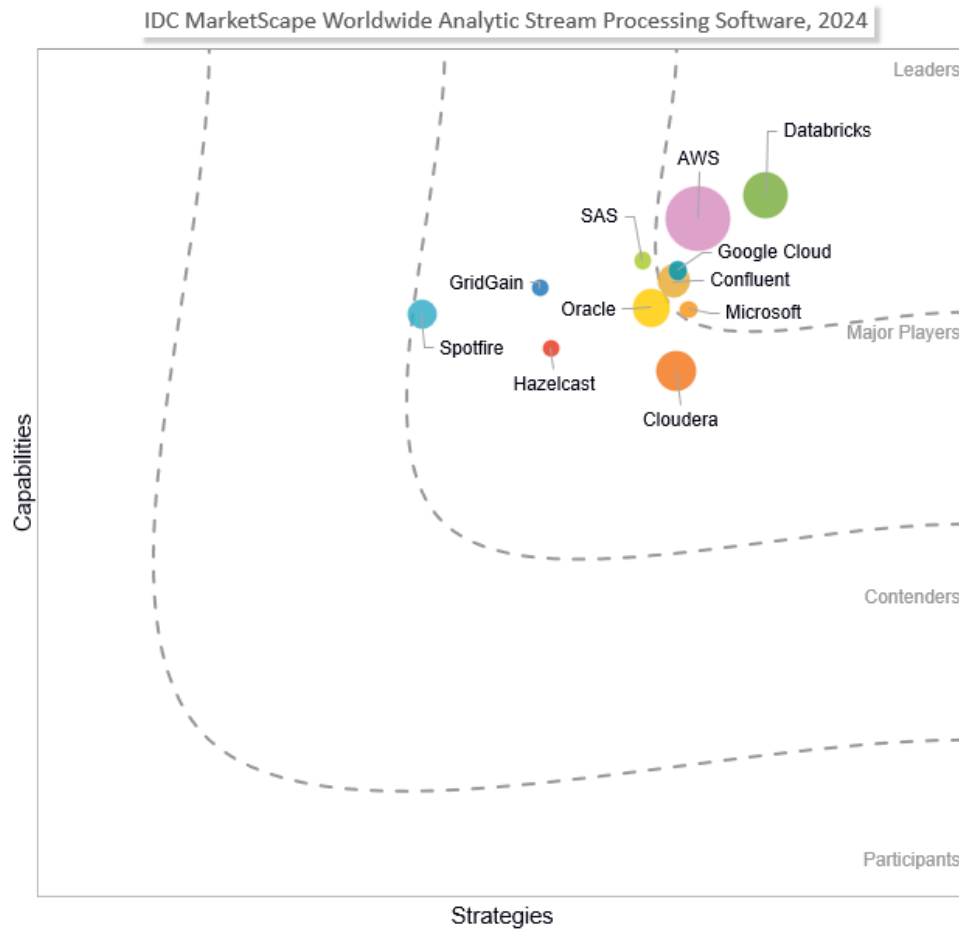
Stewart Bond

THIS IDC MARKETSCAPE EXCERPT FEATURES AWS

IDC MARKETSCAPE FIGURE

FIGURE 1

IDC MarketScape Worldwide Analytic Stream Processing Software Vendor Assessment

IDC MarketScape Worldwide Analytic Stream Processing Software, 2024

Source: IDC, 2023

Please see the Appendix for detailed methodology, market definition, and scoring criteria.

IN THIS EXCERPT

The content for this excerpt was taken directly from IDC MarketScape: Worldwide Analytic Stream Processing Software 2024 Vendor Assessment (Doc # US51053123). All or parts of the following sections are included in this excerpt: IDC Opinion, IDC MarketScape Vendor Inclusion Criteria, Essential Guidance, Vendor Summary Profile, Appendix and Learn More. Also included is Figure 1.

IDC OPINION

The speed of business has increased, as organizations need to respond to and make decisions based on what is happening now, not what happened yesterday, last week, or last month. The number of real-time analytic solutions in organizations has been increasing, as so too has the amount of money being spent in the streaming analytics software markets. The worldwide event stream processing market as measured by IDC has seen a 21% compound annual growth rate (CAGR) for the five-year period from 2017 to 2022 and a forecast 30% CAGR for the five-year period from 2022 to 2027. Streaming solutions are present in all major geographies around the world, across all major industries, but they are more prevalent within financial services. Popular use cases include financial analysis, real-time recommendations, providing service guarantees, cybersecurity, risk management, media and gaming, Internet of Things (IoT) applications, geospatial analysis, geofencing, and vehicle tracking.

Analytic stream processing software is available as a service or as client-managed implementations on global hyperscale cloud services and from independent software vendors. Not all analytic stream processing solutions are the same because there is not just one way to process events in a stream in real time:

- Processing events directly in the streaming data technologies (event brokers)
- Processing events at the source, such as capturing database change events
- Ingestion of events into high-performance in-memory data stores for processing
- Ingestion of events into a data lake (house) for processing

Each method has varying degrees of latency that are supported from nominal (less than 200ms) to ultra-real time (less than 40ms). Events can be stateful or stateless; are either pushed, pulled, or persistent; and one event at a time or in a micro batch.

Regardless of how the events are processed, the common denominator across all is time, more specifically, time windows. Analytics starts with querying data, and stream analytics requires a time boundary or window for the query. Windows can be fixed, sliding, tumbling, session based, custom defined, memory optimized, and time-stamp controlled. This represents the biggest change that a data or application engineer needs to understand when working with streaming data compared with working with at-rest data.

Analytics include event aggregation, filtering, transformation, routing, and enrichment for use in analytical models. Output from real-time streaming analytics can go into real-time dashboards, decisioning algorithms, applications, data lakes, data lakehouses, data warehouses, back into the stream, time-series databases, or artificial intelligence (AI) models. As the event brokering technologies, event types, and formats have evolved over the years, so too has the technology being used to process events.

There is a general tension in the market between proprietary stream processing engines and open source engines. Solutions using proprietary engines express performance, scalability, and direct integration with sources and targets of analytic stream processing. Open source solutions started with Apache Storm, then Apache Spark became more prominent, and now Apache Flink is becoming prominent in stream processing solutions. There are also innovations in the open source community in support of standard data formats such as Avro, JSON, and Parquet and stream programming frameworks such as Apache Beam. Most popular programming languages are supported, with SQL, Java, and Python being most prominent.

We are entering an era of AI everywhere, and the top 4 business drivers for AI as measured by IDC require real-time streaming data: improvements in quality, business agility, employee productivity, and customer experience. Analytic stream processing will become critical for success in AI. The need for real-time data is prevalent in top AI use cases by industry as measured by IDC:

- 3 of the top 4 AI use cases in the federal and local governments require real-time data
- 5 out of the top 6 AI use cases in banking and manufacturing require real-time data
- 4 out of the top 5 AI use cases in retail and utilities require real-time data

From automation of business processes to automated management of technology in the enterprise, real-time data is critical across all these use cases because business happens in real time. Batch processes were introduced because of technological limitations. The ability to process data in near real time has been available for at least two decades, but only now are we starting to see organizations transition as technology gets better at responding to what is happening in real time and trust in that technology increases.

Every vendor evaluated in this IDC MarketScape provides real-time analytic event processing and integration with AI technologies to meet the demands of running a digital business in modern times, where the speed of business is increasing and the time for organizations to respond to changes is decreasing. This IDC MarketScape evaluates vendors across stream processing capabilities and go-to-market strategies based on qualitative and quantitative market research that provided insight into what was important for customers wanting to compete as digital businesses.

IDC MARKETSCAPE VENDOR INCLUSION CRITERIA

The vendor inclusion list for this document was designed to accurately depict the vendors that are most representative of any given analytic stream processing software buyer's selection list. Vendors were then surveyed and further investigated to ensure that the offerings qualified with both capabilities and strategies related to the cloud integration market.

Critical to this research effort was for the vendor to meet the inclusion criteria. Any vendor participating in this IDC MarketScape had to showcase that it met the following:

- Offers analytic stream processing capabilities as part of a sustainable business, whether offered standalone or part of a larger technology portfolio (if part of a larger portfolio, the analytic stream processing solution must be technology agnostic – there must be no requirement to have adopted a solution from that portfolio, for example, a software-as-a-service [SaaS] application from that vendor])

- Offers analytic stream processing that fully supports analytical use cases including data ingestion and sharing, event aggregation, transformation, and filtering and can perform stream integration
- Has at least $25 million in revenue for 2022, specifically from analytic stream processing software/services offering regardless of where it is deployed
- Operates in North America and recognizes a minimum 5% of revenue from at least two additional regions
- Goes to market horizontally across industries and across horizontal processes of an enterprise

ADVICE FOR TECHNOLOGY BUYERS

Every solution acquisition initiative needs to start with the requirements for the solution. Streaming analytics can be a solution specific to one project, but it is better leveraged as a shared enterprise service capable of being used across multiple use cases and projects. As such, buyers should be considering enterprise-level requirements while looking at project-level requirements. Some requirements to consider include the following:

- What type of latency do I require for my analytic stream processing use cases –nominal (less than 200ms), real time (greater than 40ms, less than 200ms), or ultra-real time (less than 40ms)?
- Is the use of open source important? Latency requirements, how events are processed, and integration with analytics solutions will influence your decision.
- The data lake, lakehouse, and warehouse technologies being used in the enterprise may influence the decision as each solution has different alignments in the ecosystems that exist in the market.
- What types of event broker technologies are being used across the enterprise? Kafka is supported by all and most of the commonly used message-oriented middleware software. Pulsar support is growing.
- While many of the analytic stream processing software have graphical programming environments with low-code and no-code capabilities, most also support the use of commonly used programming languages. Assess programming skills within your organization against the languages supported by these technologies.
- Support of industry-specific capabilities, content, models, and so forth is sporadic among the vendor solutions evaluated in this IDC MarketScape. If having industry-specific capabilities is important, choose wisely.
- What levels of security and privacy need to be supported by the solution? Do you have FIPS or FedRamp requirements, HIPPA, GDPR, or SOC requirements?
- Governance and control are key elements of data processing in a digital business. Consider what types and level of intelligence you require about events and event processing. Catalogs, schema registries, lineage, and quality enforcement are available within the market.

Working with streaming data requires a mind and paradigm shift for developers and operations personnel; identify the levels of training you will require for developers and operations staff. Some of the open source stream processing engines can be very complex to manage, which is why most of the vendors that offer open source support make their offering available as a managed service. Consider whether a managed service is required, or will your IT organization manage the solution? What level of

monitoring and alerting is required, and does it need to integrate with existing systems management investments. What deployment options are preferred, such as cloud, on premises, or hybrid?

All vendors in this evaluation offer flexibility in licensing and pricing but to varying degrees. Consider the most optimal approach to licensing for your organization based on your use cases, technology, and budget constraints.

The evaluation of analytic stream processing software within this IDC MarketScape is based on the criteria that IDC created, informed by research. Buyers need to perform their own evaluation against their own requirements within the constraints of their own IT, procurement, and industry environmental constraints. Every evaluation process should also include proof of concept or what is often referred to as a "bake-off." These evaluations of the software need to be done against data used by your organization, in the context of your organizational structures, and within the constraints of your organization's technology environment.

Once the selection process has completed, the hard work of implementation begins. Change management will be key to successful implementation because use of stream analytics will change the way people work and how the enterprise functions.

VENDOR SUMMARY PROFILES

This section briefly explains IDC's key observations resulting in a vendor's position in the IDC MarketScape. While every vendor is evaluated against each of the criteria outlined in the Appendix, the description here provides a summary of each vendor's strengths and challenges.

AWS

After a thorough evaluation of AWS' strategies and capabilities, IDC has positioned the company in the Leaders category in this IDC MarketScape for worldwide analytic stream processing software.

Amazon Web Services (AWS) is a worldwide hyperscale cloud service provider, was founded in 2006, and is headquartered in Seattle, Washington, United States. AWS has several data-oriented services available in its cloud, including its Managed Service for Apache Flink (previously known as Amazon Kinesis Data Analytics). The Amazon Managed Service for Apache Flink makes it easy to build and run real-time stream processing applications using Apache Flink. Amazon Managed Service for Apache Flink provisions and configures Flink clusters and orchestrates Flink job management, allowing the developer to focus on higher-value functionality in streaming data pipelines. The Amazon Managed Service for Apache Flink has several companion products including Amazon Glue Streaming ETL, Amazon Glue Schema Registry, Amazon Glue Data Catalog, and Amazon EMR.

Quick facts about Amazon Web Services include:

- **Product name:** Amazon Managed Service for Apache Flink
- **Employees:** 100,000+ (estimated)
- **Global footprint:** Operates across all global and within many regional locations around the globe
- **Top industry areas:** Financial services, manufacturing, telco, media and entertainment, and high tech
- **Cloud:** Available on the AWS Cloud

- **Pricing model:** Usage based by time duration and level of compute
- **Partner ecosystem:** Global partner ecosystem through the AWS Partner Network
- **Interesting fact:** AWS has offered a Managed Apache Flink service since 2018 (formerly called Kinesis Data Analytics) and supports tens of thousands of applications.

AWS was one of the first companies to offer a managed Flink service, and because of this, the company has depth of knowledge and capabilities in operationalization of analytic stream processing using Flink. The Amazon Managed Service for Apache Flink has proven availability, security, and durability of streaming solutions using the Apache Flink open source project. The Amazon Managed Service for Apache Flink is serverless, it supports all Flink APIs, and customers can use Java, Python, Scala, and SQL to create their applications. It supports Flink's exactly once processing semantics and state management configs, and it has over 40 prebuilt connectors and supports over 200+ custom connectors. The Amazon Managed Service for Apache Flink Studio is an interactive experience for stream querying and promoting queries into long-running Flink apps. AWS supports customers' access to specialist Flink engineering, solution architects, and support, including Flink PMC members and committers.

AI-assisted development is an active and growing opportunity for data stream processing that AWS is pursuing, anchored by Amazon Q, which was announced at AWS re:Invent in the fall of 2023. Amazon Q is a generative AI-powered solution that provides opportunities for automation of processes from requirements gathering through development, deployment, and ongoing management of solutions running on the AWS Cloud.

Strengths

- **Compatibility and language support.** The Amazon Managed Service for Apache Flink supports multiple types of event brokers such as Kafka, Pulsar, RabbitMQ, and MQTT. It also supports all popular programming languages including C++, C#, Java, Python, SQL, Scala, Ruby, Julia, Go, and R. It also supports the use of Apache Beam and Spark.
- **Market presence and partner ecosystem.** The Amazon Managed Service for Apache Flink is part of the AWS Cloud, which offers availability across all global regions, providing sales and service support in 53 countries. AWS has a vast partner network inclusive of channels, services, and technology companies. Early in 2023, AWS launched a partner program explicitly focused on the Amazon Managed Service for Apache Flink.

Challenges

- **Awareness.** The Amazon Managed Service for Apache Flink was formerly named Kinesis Data Analytics. The market did not know that Flink was behind the Kinesis Data Analytics service, and AWS was missing the opportunity as Flink grew in popularity for stream processing. This is the premise behind the recent rebranding into the Amazon Managed Service for Apache Flink. Apache Flink requires a steep learning curve to build and operate, and AWS saw an opportunity to be more transparent and better align with growing industry trends.
- **Cloud deployment alternatives.** The Amazon Managed Service for Apache Flink is only available on the AWS Cloud. This, however, does not preclude it from being able to connect to event brokering technologies running in hybrid and multicloud environments using secure connections.

Consider AWS When

Consider AWS when your organization has a disposition toward AWS as a cloud service provider and, as such, much of your streaming data infrastructure is also contained within the AWS Cloud. You may also consider AWS if you are using Flink as a stream processing engine and want to leverage a managed service to accelerate the learning curve and/or reduce complexity of operations, managing Flink clusters, OSS patches, and maintenance.

APPENDIX

Reading an IDC MarketScape Graph

For the purposes of this analysis, IDC divided potential key measures for success into two primary categories: capabilities and strategies.

Positioning on the y-axis reflects the vendor's current capabilities and menu of services and how well aligned the vendor is to customer needs. The capabilities category focuses on the capabilities of the company and product today, here and now. Under this category, IDC analysts will look at how well a vendor is building/delivering capabilities that enable it to execute its chosen strategy in the market.

Positioning on the x-axis, or strategies axis, indicates how well the vendor's future strategy aligns with what customers will require in three to five years. The strategies category focuses on high-level decisions and underlying assumptions about offerings, customer segments, and business and go-to-market plans for the next three to five years.

The size of the individual vendor markers in the IDC MarketScape represents the market share of each individual vendor within the specific market segment being assessed. For this IDC MarketScape, vendor size was determined using IDC's 2022 Software Tracker, which represents IDC's estimate of each vendor's 2022 software product revenue.

IDC MarketScape Methodology

IDC MarketScape criteria selection, weightings, and vendor scores represent well-researched IDC judgment about the market and specific vendors. IDC analysts tailor the range of standard characteristics by which vendors are measured through structured discussions, surveys, and interviews with market leaders, participants, and end users. Market weightings are based on user interviews, buyer surveys, and the input of IDC experts in each market. IDC analysts base individual vendor scores, and ultimately vendor positions on the IDC MarketScape, on detailed surveys and interviews with the vendors, publicly available information, and end-user experiences in an effort to provide an accurate and consistent assessment of each vendor's characteristics, behavior, and capability.

Market Definition

Analytic stream processing software ingests and processes events by normalizing the events into a data structure, correlating the data to the correct stream, applying time window-based logic to evaluate the data, querying additional data to apply business logic, and may include the use of algorithms to apply advanced analytical models to the event stream and routing or delivering the output to business logic that determines what to do with the output. Stream processing supports a wide range of use cases that support automation and automated decision-making, including descriptive and predictive analytics.

Related Research

- *IDC MarketScape: Worldwide Event Brokering Software 2024 Vendor Assessment* (IDC #US50301923, March 2024)
- *IDC PeerScape: Practices to Improve Analytic Stream Processing* (IDC #US51712624, March 2024)
- *IDC Market Glance: Data Control Plane, 1Q24* (IDC #US51070924, February 2024)
- *Modern Data Streaming* (IDC #US49529023, August 2023)
- *IDC PlanScape: Future of Intelligence – Data Engineering* (IDC #US47348321, June 2023)

Synopsis

This IDC MarketScape on worldwide analytic stream processing software vendor assessment for 2024 discusses the increasing need for real-time analytic solutions in businesses and growth of the streaming analytics software market. This document evaluates various vendors based on their stream processing capabilities and market strategies. It also provides advice for technology buyers and outlines the vendor inclusion criteria. The document concludes with a summary of each vendor's strengths, challenges, and considerations for potential buyers.

"The speed of business has increased, as organizations need to respond to and make decisions based on what is happening now, not what happened yesterday, last week, or last month, " said Stewart Bond, vice president, Data Intelligence and Integration Software Research at IDC. "Organizations require real-time analytics solutions fed by streams of data; the market has responded and is experiencing significant growth, but not all analytic stream processing products are created equal, thus requiring an objective evaluation of what is in the market."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology, IT benchmarking and sourcing, and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.).

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com