Bloor

Market Report

Market Report Paper by Bloor
Author **Ronnie Beggs**
Publish date **June 2016**

Streaming analytics 2016

The future of streaming analytics lies at the heart of the real-time enterprise for driving business process automation.

Author **Ronnie Beggs**

Streaming Analytics 2016

Where capital markets laid the foundations and Big Data planted the seeds for growth, the Internet of Things is driving wider industry adoption and propelling streaming analytics into the mainstream. Not that it's all about sensors. The vendors in this report demonstrate concrete and significant deployments across a range of other industries. Telecoms, retail, capital markets and financial services feature prominently, with substantial deployments for horizontal solutions such as customer experience, real-time promotions and fraud detection. However, it is the Industrial Internet of Things (IIoT) and M2M where we are seeing a rapid increase in vendor and market activity, with significant deployments of streaming analytics in sectors such as healthcare, smart cities, smart energy, industrial automation, oil and gas, logistics and transportation.

The raison d'être for streaming analytics is to extract actionable, context-aware insights from streaming data, the type of data emitted by devices, sensors, web platforms and networks for example, and to help businesses respond to these insights appropriately and in real-time. However, it is becoming apparent that streaming analytics merges traditional IT and big data with the world of operational technology. For operational users, visualization for real-time monitoring is important, but the term 'real-time' often implies automated actions and end-to-end process integration. As we discuss in this report, the future for streaming analytics is more than merely a conduit for real-time analytics to business applications. The future of streaming analytics lies at the heart of the real-time enterprise for driving business process automation.

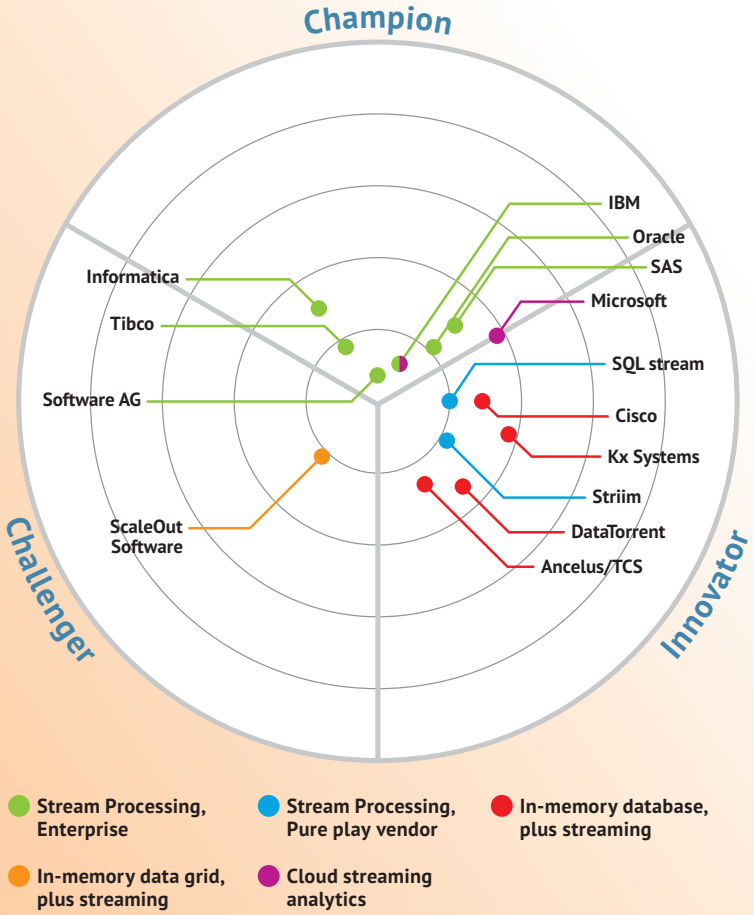
The Vendor Landscape

There is more than one way to crack a nut. Perceived wisdom has associated streaming analytics with data stream processing. Yet enterprise requirements vary, and not surprisingly, our research indicates that businesses are adopting a range of approaches, including fast in-memory databases with streaming data extensions, and even in-memory data grids for the event-driven analysis of streaming data. We have therefore broadened the report to include the wider range of technologies for delivering streaming analytics that we are seeing in enterprise environments.

Included vendors: Cisco, DataTorrent, IBM, Informatica, Kx Systems, Microsoft, Oracle, ScaleOut Software, SAS, Software AG, SQLstream, Striim, TIBCO Software, Time Compression Strategies.

"The raison d'être for streaming analytics is to extract actionable, context-aware insights from streaming data, and to help businesses respond to these insights appropriately and in real-time."

Figure 1: The Vendor Landscape



Our key observations of the Streaming Analytics Market in 2016

> The Industrial Internet of Things (IIoT) requires hybrid, hierarchical cloud architectures with intelligent appliances at the edge of the network. Streaming analytics is an essential component of edge intelligence.

- **To stream or not to stream is no longer the question.** Streaming analytics is here to stay. The market is demonstrating significant growth and is fulfilling its potential. The number, diversity and sophistication of the customer deployments exhibited by the vendors in this report bode well for the future. The maturity and stability of the products is impressive, with customer installations exhibiting compelling business and operational benefits, and deployments that have been in production for many years.
- **Better tools for application development means better productivity.** Where streaming analytics was once the realm of the developer, there is now a trend towards engaging directly with the business analyst community through graphical tools for rapid application development without coding. The use of solution libraries for accelerating application deployment is on the up, strengthening differentiation against the open source stream processing frameworks, and indicating the emergence of a more value-oriented enterprise approach.
- **Predicting is the future.** Automating preventative actions in the real world requires a high level of confidence in the analytics, and in general, a level of sophistication and context that is beyond the capability of standalone streaming analytics solutions. There is a clear trend towards integrating streaming analytics with predictive analytics, machine learning and rules engines, in order to improve accuracy and reliability, and to simplify the expression of complex business patterns and scenarios.
- **Real-time without process automation is really human time.** Streaming analytics blurs the boundary between business intelligence and operational processes. The industry is looking beyond real-time monitoring, operational intelligence and real-time dashboards, to the full-blooded automation of corrective actions. Increasingly, vendors are partnering or utilizing their wider product portfolios to deliver complete enterprise-class architectures for real-time systems. This requires the integration of streaming and predictive analytics with operational process management platforms.
- **Streaming analytics, a shining light in the Fog.** The Industrial Internet of Things requires hybrid, hierarchical cloud architectures with intelligent appliances (and applications) at the edge of the network. Streaming analytics is an essential component of edge intelligence, and vendors are now looking at embedded and appliance-based deployments as a complement to their cloud solutions. There is evidence that advanced features are being considered, if not as yet fully addressed, including remote application management at scale, intelligent data synchronization over poor quality communication links, and how the appliance behaves on power up after the plug gets pulled.

The need for speed. Myth or reality?

The customer references taken up as part of this report made it clear that streaming analytics platforms deliver a level of performance that cannot be achieved at any cost by traditional approaches to data management, even Hadoop. The majority of references reported low latency as a key requirement – the ability to deliver results in milliseconds up to a few seconds as measured from the time of data arrival. Of course, one could argue that latency of a few seconds does not require streaming analytics. However, solution latency is more often than not a factor of the operational process within which the system is working, and in all cases, continuous processing was key to success.

However, our research suggests that throughput performance is not a significant differentiator, although adequate performance and an assurance of future scalability are essential. This is not to deny that there are use cases where the ability to process millions of records per second is top of the customer's priority list. However, the wider market is more conservative in its performance requirements, but less tolerant of incomplete, unstable solutions.

In fact, evidence suggests that the answer lies in the package of performance, resilience, scalability, ease of development, stability and ease of integration, plus the ability to deliver (or integrate with) the wider portfolio of products required to deliver complete real-time systems.

Streaming analytics platforms deliver a level of performance that cannot be achieved at any cost by traditional approaches to data management.

Where is the disruption?

The combination of capital markets, financial services, and banking remains the largest overall market for streaming analytics in terms of revenue. This may be a disappointing observation for some, particularly those hoping for excitement elsewhere. It is certainly not disappointing for those with a presence in these sectors, particularly with new use cases emerging for fraud detection, consumer experience, and consumer products for context-aware analytics services.

It could be argued that the Complex Event Processing (CEP) vendors were successful in capital markets and financial services by disrupting an existing IT market, offering flexible platforms, rapid deployment and high performance in a market typified by bespoke, complex (pun intended) and expensive systems. One could take this further and suggest that rather than CEP failing to break out into other industry sectors, that this was the only industry sector for some time where disruption was evident.

If we were to observe the list of target markets beyond the triumvirate of capital markets, financial services and banking, there would be a strong correlation between a list of use cases from any vendor's website or market report from say ten years ago, with a list of use cases today. Telecoms, ad promotions, consumer experience, website visitor tracking, real-time pricing would be present in both. What is changing is at last there are signs of growth with substantial, revenue-generating deployments in all these markets. The majority of vendors have customers in telecoms for CDR processing for example, and most have customers for customer experience management, fraud detection and retail inventory management or real-time promotions. But why has it taken ten years?

Ironically, the business case streaming analytics would have been much simpler and clear-cut had Hadoop not happened. The shortcomings of the RDBMS as a platform for real-time, operational analytics from unstructured event data would have been an easy target. Hadoop offered to reduce business latency from one day to a few hours or less, which as it turned out was sufficiently compelling for many, even if not real-time.

Our research also suggests that the Industrial Internet of Things (IIoT) and connected devices (M2M) have the potential to topple financial services as the leading market. The characteristics of streaming analytics are particularly suited to the processing of sensor data: the combination of time-based and location-based data analysis in real-time over short time windows, the ability to filter, aggregate and transform live data, and to do so across a range of platforms from small edge appliances to distributed, fault-tolerant cloud clusters. Sensor data volumes have already reached a level where streaming analytics is a necessity, not an option.

However, as with any market, a step change in adoption of IIoT services requires real disruption. In one direction we have the traditional IT-oriented streaming analytics market, with a focus on visualization and generating actionable insights, taking data streams from sensors and transforming these into alerts, graphs and reports. This is a valid market of course, as any emerging operational market requires real-time performance and event management. Compare and contrast for example with the telecoms industry, the rise of cellular wireless technology, and the need for real-time network monitoring systems. That said, this is a very active area with significant startup activity from different directions – cloud PaaS, operational intelligence products, and existing IoT, sensor and device manufacturers. Therefore, market positioning and go-to-market focus for a streaming analytics vendor in this sector is key.

Beyond this IT-oriented market for operational intelligence lies a world inhabited by management and control systems, hardware and devices. This is an area that remains largely untouched by cloud and big data technology, and which covers many industries including retail, transportation, logistics, energy and industrial automation. True disruption for IIoT means the replacement of traditional monitoring and control applications with a combination of cloud-based software platforms, virtualized infrastructure and edge appliances.

For many organisations, this is the replacement of its entire operational management infrastructure – from manual processes, through to their real-time monitoring systems, data historians, inventory management and control systems. This requires a focus on business transformation, and the integration of sensor data with business process. Even RFID monitoring, possibly the most-quoted of all use cases for streaming analytics, is finding a home as RFID data is used to drive real-time inventory, stock management and pricing.

We believe those vendors best placed to exploit the Internet of Things are those who offer a broad product portfolio combined with solutions services and consulting. Or for smaller vendors, those who identify correctly their position in the overall value chain and align their products and company accordingly.

Markets and use cases

Significant business value is being derived from streaming analytics in many areas. The following market reviews consider feedback from the vendors included in this report, working on the assumption that a survey of the leading vendors is as good an indication as any as to the current state of the market. However, we have also considered the growth potential and competitive pressures going forward.

The Internet of Things

- **Preventative maintenance** could well be the silver bullet for streaming analytics in IIoT. We have seen preventative maintenance emerge as the leading use case in this sector, and we believe the one with the greatest potential. Several vendors cited preventative maintenance customers across different markets, including vehicle telematics, oil and gas drilling equipment on remote rigs, conveyor belt wear, elevators, and pipeline leaks. The value to the customer is clear, to reduce operational and equipment cost by minimising unplanned outages, and to reduce the requirement for expensive site and maintenance visits that could be avoided.
- **Retail.** RFID sensors have at last found a useful existence. Of course, it's not all about RFID technology, but real-time inventory updates are helping to drive business processes for inventory and pricing optimisation, and for optimisation of the supply chain, logistics and just-in-time delivery. This is also a market where wearables and the consumer market is poised for even greater growth. The wrist may not be the best place for a bunch of sensors. Yet, in a market obsessed with timepiece replacement for those with deep pockets, it could be the garment as a sensor that is likely to offer greater revenue potential. Streaming analytics will have a role to play in use cases including healthcare and fitness.
- **Smart Transportation.** Uber has shown where true market disruption can lead (Uber makes extensive use of stream processing and streaming analytics). The model in transportation is now set towards usage-based pricing and operations. We reviewed use cases including telematics for usage and location-based insurance (although this is already a market with strong startup activity from vertical application vendors), and driver advisory systems with feedback to the driver to cut petrol and diesel costs. The future is certain to bring more use cases for usage based pricing, for example, tyre manufacturers are experimenting with smart sensors in tyres to measure usage and wear, and smart city car share schemes that combine usage-based pricing models with real-time tracking and vehicle telematics.
- **Smart Energy.** We found several deployments in the smart energy sector, from real-time monitoring of smart meters, smart pricing models for electricity, to real-time sensor monitoring of wind farms (which produce a vast volume of sensor data and where streaming analytics can drive a significant increase in efficiency and energy output). This is a new and attractive market with tangible business benefits for streaming analytics, but one where it is difficult to see where true disruption can occur. Therefore a valid market, but not necessarily one with significant growth potential.
- **Industrial automation** combines streaming and predictive analytics to optimize manufacturing processes and product quality. Where companies have implemented six sigma and lean manufacturing techniques, streaming analytics enables statistical analysis of the manufacturing process, with alerting and automated shutdown when quality levels are breached. This is an obvious area for streaming analytics, but it is also highly specialised and technical.

> **"Preventative maintenance has emerged as a leading use case for IIoT, and we believe the one with the greatest potential."**

- **Healthcare.** M2M services for improving client engagement have been around for many years without significant uptake, probably due to the shortcomings of SMS as a delivery mechanism. This is a large market offering significant opportunity and crying out for disruption. Smart sensors may unlock the potential. For example, where an SMS message can only remind a patient to take a pill, a smart sensor on a pill bottle can report continuously if a pill has been taken and when, even if the storage temperature is not correct.

Other markets:

- **Capital Markets** remains a strong sector for those with a CEP heritage. This is not about whether or not CEP is different from stream processing, that argument has long since disappeared as CEP platforms have evolved, rather it's about a strong customer base, expertise and market-specific capability. We saw several use cases for automated, algorithmic trading and for real-time trade compliance and audit, as you would expect, but also an increase in deployments for fraud detection and trading analytics as a service. The latter is the aggregation of real-time data from multiple internal sources into context-aware analytics streams, with delivery channels for both business partners and individual consumers.
- **Telecoms.** Market growth surveys for the telecoms sector should be taken with a pinch of salt. Telecoms networks are the carriers for both the Internet and the Internet of Things, which can mean double accounting for data volume and growth potential. And of course, all major service providers have deployed their own IoT platforms and services. We reviewed use cases in three broad areas: processing call detail records (CDRs), cell and IP network performance monitoring, and customer experience monitoring. Again, valid use cases addressing real customer needs, but not disruptive. We were hoping to see movement from more vendors into emerging areas such as NFV and SDN.

Horizontal markets

This section examines the main horizontal solutions that are evident across multiple industries. These offer valid markets in their own right for vendors wishing to specialize, although in general, most vendors in the report have customers across all sectors.

- **Real-time monitoring** of networks, equipment and devices is a common use of streaming analytics. Examples include wireless network monitoring for dropped call analysis, IP network monitoring for capacity and routing issues, wind farm sensor monitoring, real-time emergency call analytics, traffic congestion from GPS sensors, and railways (monitoring trackside and on-board sensors with GPS position data). However, this is also an area with significant competition from vertical application vendors, where the opportunity to be disruptive is low, and a horizontal solution area that will become increasingly competitive across all vertical industries.
- **Fraud detection.** This appears to be a sector ripe for growth given the large discrepancy between use case examples discussed and actual use case deployments in operation. It would seem that reality has not yet caught up with hype. We reviewed some strong references for streaming analytics in banking and transaction fraud in particular. However, despite there being a clear business case, the fraud and wider security, surveillance and even cybersecurity markets appear to be offering potential rather than immediate revenue opportunities. Competitive pressure from vertical application vendors in this sector is a contributing factor, who may offer domain expertise, but at the expense of the flexibility and scalability offered by streaming analytics platforms.
- **Real-time promotions** and location-based services tend to go together. Broadly, use cases spilt into clickstream and visitor analytics, plus consumer location tracking and related services. This is a sector that requires big data scalability, temporal and spatial analytics, predictive analytics, and sophisticated use of windows-based processing such as sessionization. It plays well to the strengths of streaming analytics, offering disruptive solutions that are difficult to deploy using other technologies. We expect this market to offer significant potential for those able to deploy complete solutions for real-time operational processes.
- Requirements for **customer experience monitoring** differ significantly across industries. This is an area of some hype, yet few real examples, in large part down to organisations having more powerful and blunter tools at their disposal such as pricing models. However, we were pleased to see some significant examples of streaming analytics in this area, for example, improving the customer call center experience and improving the success of incoming sales calls, both through assisted operations and real-time access to context-aware information and sentiment analysis.

The essentials

The vendors selected in this report all offer sophisticated platforms although feature sets vary. Of course, this is a diverse market where the leaders in any report may not necessarily be the best fit for a specific use case. There are however a number of key attributes of a streaming analytics solution that are useful to bear in mind when selecting a product vendor. Notwithstanding business drivers such as productivity and cost, the scale of the overall project, and the extent to which IT skills are available in-house, the following technical criteria should be considered:

- **Stream granularity.** The majority of streaming analytics platforms built on data stream processing engines offer record-by-record processing and are therefore able to react to each record as it arrives. This is essential for threshold detection for example, or fine-grained pattern matching. Most systems also support micro-batching, where records are processed in sets of fixed-duration time windows, typically to increase throughput performance and ease of aggregation prior to dataload. However, micro-batching is not suitable where the creation time and value of each record is important.
- **Time.** Time seems an obvious criterion to support, yet the level of native understanding of time varies. In most applications, it is essential to process by the time of data creation rather than the time of data arrival (unless of course the creation time is not available, in which case the system should default to arrival time). Streaming analytics operations must operate over time windows, as a minimum, supporting sliding windows for streaming analytics, and tumbling windows for data aggregation. The ability to manage time-related issues such as missing data or out of time order data requires a native understanding of timestamps and time-aware processing.
- **Location.** Location-based services and geospatial analytics are an important weapon in the streaming analytics armory. Important features include geo-fencing, nearest neighbour, and spatial measurements for area, distance and volume. Geospatial operators may be included natively, or integrated using an external library, why re-invent the wheel. Consider the impact of geospatial operations on performance, and the extent to which an external library has been adapted for real-time stream processing rather than static data analysis.
- **Context.** Real-time data streams rarely contain sufficient context on which to base operational decisions, for example, customer details will not be available in data originating from a telecoms network. Enrichment enables live data to be enhanced by pulling in reference data from external databases, storage platforms and in-memory data grids. Most vendors support the enrichment of streaming data, however not all offer caching of static data, particularly in distributed environments. System performance can be impacted as a result.
- **Distributed architecture.** All vendors exhibited excellent system performance for throughput and latency over multiple CPU cores on a single server. Given their highly optimized architectures, performance on a single server is sufficient for many applications. However, support for distributed processing varies, in particular, not all architectures support distributed processing, fault tolerance and resilience without application level coding.
- **Operational management.** All vendors have experience in the deployment and upgrade of operational systems, although support varies for the implementation, test, approval and deploy process. Facilities for dynamic updates to operational systems should be reviewed carefully, particularly in mission critical environments where downtime is a major issue. Support for dynamic updates and remote management varies but how important this is depends entirely on the application.

Facilities for dynamic updates to operational systems should be reviewed carefully, particularly in mission critical environments where downtime is a major issue.

A progression from data and technology to applications and process

The streaming analytics market is wide and diverse, covering numerous vertical industries and horizontal solutions. We have opted to classify the market based on the level of business value extracted from the data, and have proposed four categories: continuous dataload, real-time monitoring, predictive analytics, and real-time automated processes. This is by no means a rigid scale, nor is it necessarily incremental. Some vendors have the capability to address all four levels in a single installation. Others focus on a single level. For example, many streaming analytics deployments focus on continuous integration or real-time monitoring, but some, particularly those in industrial automation, may operate solely in black box, lights-out mode for process automation and control. We have included the product capabilities that we would expect to see at each of the levels.

Level 1: Filling the Data Lake.

Essentially, real-time data warehousing, with the collection, continuous aggregation and dataload of streaming data into the Hadoop data lake or RDBMS data warehouse. Streaming analytics platforms with graphical development tools deliver a compelling and feature-rich solution. However, open source projects offer an alternative for simple data ingest into Hadoop, and existing ETL vendors are offering Hadoop ingest using the same open source projects. Therefore streaming analytics vendors tend to position data stream ingest as the building block for operational intelligence and process automation.

Key attributes:
Tumbling windows with aggregation and filtering operations, a comprehensive range of connectors for machine data capture, Change Data Capture (CDC) for databases, Hadoop and database adapters, enrichment operators, graphical development tools supporting end-to-end streaming ETL pipelines.

Level 2: Real-time Monitoring.

The term operational intelligence is also valid in this context, but as a market, operational intelligence includes a much wider range of technologies and approaches. Streaming analytics platforms for operational intelligence deliver better performance, scalability, flexibility and interoperability with enterprise systems. Visualization is a key requirement. All vendors in this report package a visualization platform for streaming analytics, or offer integration with third party BI dashboard products. The general level of capability across the vendors is good, with some mature product offerings.

Key additional attributes:
A rich set of streaming analytics operators including support for geospatial analytics, integrated and user-configurable dashboards, ability to push context-aware streaming analytics across multiple delivery channels (API data feeds, mobile, web, device for example).

Level 3: Predictive analytics driving human intervention.

The ability to prevent and optimize responses to future failure scenarios enhances the business case for streaming analytics considerably. Predictive analytics can be developed using the development tools provided by any of the platforms, but this is where those with integrated rules engines and predictive analytics start to shine.

The delineation of human responsibility from machine responsibility is important. Operational process automation requires change management and business process integration, the scale of which increases as systems move towards the mission critical.

Key additional attributes:
Integrated predictive analytics, rules engines, with sophisticated alerting and pattern matching capability.

Level 4: Operational process automation.

You have to be certain if you're going to halt a production line, shut down a drilling operation, automate a stock trade, even to update a pricing model based on real-time inventory information. This level of sophistication requires rules engines, predictive analytics, and importantly, interaction with process management systems. This area is growing to be a significant market for streaming analytics, incorporating IIoT, banking, retail, e-Commerce, supply chain and logistics, to name but a few. Vendors with integrated enterprise product offerings and the ability to deliver business transformation projects will fare best.

Key additional attributes:
Integrated business process management, business consulting services, business transformation expertise.

Vendor selection and evaluation

Vendor Selection

We have not attempted to analyse every streaming analytics vendor in the market, and certainly not every vendor with aspirations to extend their real-time and in-memory offering into the streaming analytics arena. The report contains a representative set of the leading commercial vendors, based on our judgment and recommendations from Bloor Research subscribers. Stream processing vendors are predominant, but we have also included a number of vendors with different underlying technology where this is being used and positioned in the market for streaming analytics. We see the market as divided by technology, market focus and maturity as follows:

- A **data stream processing engine** is the popular choice on which to build streaming analytics solutions. We classify stream processing as the ability to process multiple data streams arriving at different rates on a record-by-record basis, as the records arrive. Processing operations are expressed using a range of language techniques, from ANSI SQL, through proprietary SQL-like languages, to low level development in languages such as Java, C# and C++. However, the programming mechanism is less important than a vendor's ability to identify their target buyer, set expectations correctly, and provide a rich and appropriate set of tools and development processes.
- **Fast databases** with extensions for processing data streams are popular platforms for streaming analytics. These are typically in-memory/hybrid columnar databases, and while they may not offer the range of pure stream processing operations over the live streams, those that play in this sector are able to deliver low latency and high throughput, with the advantage of data persistence as part of the package.
- The third option to emerge is the use of **in-memory data grids** (IMDG), bringing an object-based perspective to streaming analytics. Essentially an event-driven architecture, in-memory data grids offer fast and powerful processing of all data in the enterprise on a single platform, streaming and static, with the added advantage of sophisticated data synchronization capability across multiple underlying storage platforms. This is an interesting development given the wider market adoption of IMDGs. A point of note, that not all IMDG or in-memory fast database vendors position their products for streaming analytics.
- **Enterprise vendors versus pure play streaming analytics.** Enterprise software vendors with broader product portfolios are leveraging their assets to good effect. In key markets where predictive analytics and real-time process automation are required, the ability to deliver an end-end business solution on an integrated, real-time enterprise architecture is a significant differentiator. Moreover, several of these vendors also offer their own solution services and consultancy expertise. We believe expert services is key in any emerging IT market, including streaming analytics, where the customer IT and operational departments may not yet have the knowledge and skills required. That does not mean that pure play vendors are disadvantaged, but for them, strategic partnerships, go to market and positioning become critical.

Vendor Evaluation Inputs

The following inputs were required from each of the vendors participating in this report.

- **Evaluation criteria.** The vendors were required to answer a questionnaire containing 40 criteria for assessment, covering streaming analytics capability, product packaging, integration, operational management and upgrade of systems, and application development.
- **Vendor briefings and interviews.** Briefings and interviews were undertaken to offer the vendor

> **"This report contains a representative set of the leading commercial vendors, based on our judgment and recommendations from Bloor Research subscribers."**

an opportunity to explain their differentiators, market positioning, case studies, and of course roadmaps and future plans.

- **Customer references.** The participating vendors provided at least two customer references that had agreed to speak with Bloor Research. The references provided details of their implementation, including performance metrics, the type of end users and how those users accessed the analytics, and the overall business benefit of the solution to their organisation.

What were we looking for?

This is not a report looking at which vendor has the best data stream processing technology. We believe the market has moved beyond the purely technical, but is also varied in its approach. The following lists the criteria against which we scored the vendors within their stated market and go-to-market approach. Not all criteria were relevant for all the vendors and this was taken into account. However, we would like to state clearly that all the vendors have strengths in certain areas. We have explored the various market dynamics within this report with the intention of offering a framework to select the most appropriate vendor for a given set of functional and operational requirements.

- **Streaming analytics features.** The richness of higher order data processing operators for transforming data streams into streaming analytics. In particular, support for time windows, transformation, data aggregation and pattern matching, as well as geospatial and location-based capability.
- **Solution footprint.** The extent to which vendors are utilizing their wider product portfolios or partner channels to deliver end-to-end streaming analytics solutions. This includes components such as visualization, rules engines, predictive analytics, middleware and enterprise integration, and business process management.
- **Productivity tools and ease of application development.** Given the vendor's stated market and implementation focus (developer or analyst for example), an assessment of the programming language, development and test tools, including the process of transitioning changes into operational use, plus use of other libraries and solution accelerators.
- **Platform performance and scalability.** An assessment of the platform's performance in terms of throughput and latency, including scalability (scale-up and scale-out), resilience and fault tolerance.
- **Edge Processing for the Internet of Things.** Where the Internet of Things was stated as a strategic market, and where target markets are likely to require edge intelligence, an assessment of the vendor's vision for the remote deployment and management of edge applications.

The Bullseye Methodology

Bullseye is an open standard methodology for the comparative analysis of IT vendors and products that takes into account both technology and business attributes. The Landscape Bullseye is a representation of the vendor landscape split into three sectors, Champions, Innovators and Challengers, with the highest scoring companies in each sector nearest to the centre.

- Champions are generally larger organisations with a strong overall product portfolio and wider solution offering.
- Innovators tend to be smaller companies, or new products from larger companies that are strong across the market at the product level.
- Challengers offer valid products, and tend to be new entrants with potentially disruptive approaches, or companies with a specific focus within the wider market.

A few vendors have multiple classifications. Although we reviewed Cisco's streaming analytics offering, Cisco Streaming Analytics (CSA), we felt that the Parstream acquisition was a complementary component of their edge strategy. IBM also offers Streams as on-premise and on Bluemix. Microsoft continue to offer StreamInsight as an on-premise deployment although for this report the focus was on their Azure Stream Analytics.

We also felt it useful to break out enterprise software vendors as a separate category. These vendors offer complete solutions over and above streaming analytics, and typically have solutions consulting and project services. As such their go to market and target audience differs from that of the pure play product vendors, and tend to compete more with each other than the smaller vendors.

It should be noted however that the Bullseye chart is merely a starting point. Although the feature sets, strengths and market focus vary, all the vendors included in this report offer solutions that are delivering significant business benefit to their customers. As it is important to understand the positioning and strength of the various offerings, the following sections provide an overview of the vendors and products included in our research.

We strongly recommend reviewing the vendor reports in detail in order to understand the strengths and offerings of each in a wider context.

Vendors

Cisco

Cisco has an advantage in being Cisco, eliminating any doubt as to the market positioning of its streaming analytics offering. This is edge analytics for the Internet of Things, or in Cisco parlance, the Fog. Cisco's IoT architecture consists of Fog Nodes, remote appliances for sensor monitoring and data collection, with embedded edge analytics in the ruggedized edge routers and switches to which the remote Fog Nodes are connected. The architecture reflects the need for streaming analytics at the edge, but also that edge persistence of data is useful, for example, when sensors generate more data than available communication bandwidth may allow.

The core offering is built on two acquisitions, Cisco Streaming Analytics (CSA), previously known as Truviso and acquired in 2012, and ParStream, a recent acquisition, and a high performance, in-memory/hybrid disk columnar database. ParStream is optimized for fast data import with a small footprint suitable for embedded appliances. The remote device appliances, or Fog Nodes, are delivered with ParStream enabled as an embedded real-time Historian. CSA is embedded within edge routers and switches, providing in-memory processing of the sensor data streams delivered from the Fog Nodes (or directly from sensors). CSA can monitor Netflow and other infrastructure telemetry data and used for in-network performance and quality of service analytics for hyper-distributed enterprises.

CSA is best described as a data stream engine, built on the PostgreSQL database engine with extensions for continuous SQL queries over arriving data streams. CSA offers the expected range of streaming operators, with continuous queries for event by event stream processing, support for data and system time, tumbling and sliding windows, stream and table joins etc, plus native support for geospatial queries using PostGIS. As a SQL engine, operator extensions and libraries can be added as UDFs and custom aggregates as UDAs. Although high availability and fault tolerance is supported, true distributed scalability is likely to be an issue due to the age old database issue of managing state in a distributed environment. That said, this is not an issue for embedded edge device applications that tend to be single node, and where CSA's patented architecture, data compression and performance will be the driving factor.

A unique feature we believe is support for advanced cross stream semantically-based window services using a technique called sessionization. A session is a dynamic window where the window duration is determined by rules, such the start and end times of a visitor connection to a public Wi-Fi router. Useful for crowd analytics for example, executing queries across all active sessions concurrently at a sports event for example. Other vendors may provide the basic capability for dynamic window duration for example, but have not as yet packaged it in this manner.

DataTorrent

DataTorrent was launched in 2012 by founders with relevant experience from Yahoo, and has set its stall out fairly and squarely as the distributed streaming analytics platform to beat. DataTorrent offers the levels of distributed processing that we've come to expect of the Apache open source projects, but continues to add the features expected of an enterprise-level product. It recently submitted its core RTS engine as aproject under the open source Apache licence. Whilst it may be difficult to gain traction as an open source stream processing project given the competition within the Apache ecosystem, there is no doubting that DataTorrent takes the initiative seriously, and that it can only cement their credentials as the leading streaming analytics vendor on Hadoop.

Indeed, DataTorrent's reference customers listed scalability, self-healing and fault tolerance as plus points for its distributed architecture. The DataTorrent RTS engine, now Apache Apex, is an in-memory YARN-native platform that scales out dynamically over Hadoop clusters. Additional resources are added from the Hadoop cluster as processing load increases, with additional nodes utilised as these are added to the cluster. This enables Apex to scale as throughput rate increases and to handle unexpected spikes in data volume, while maintaining very low

The volume of sensor data from the Internet of Things has already reached a level where streaming analytics is a necessity, not an option.

latency responses. Fault tolerance with zero data loss is supported natively in the event of server outages or network failure, that is, no additional application level coding is required.

DataTorrent's support for streaming analytics operators, data collection and enterprise integration, is provided through packaging of the Apache Malhar library with its core RTS package. Malhar provides the standard data management operators, including filtering, aggregation, joins and transformations. Although the processing of data streams is based on a micro-batch approach, DataTorrent offers additional capability for event by event processing within the window, where windows can be as short as a few milliseconds.

The Community Edition (CE) of DataTorrent RTS includes a graphical platform for distributed platform management, with the Enterprise Edition (EE) also including dtDashboard for visualization. dtDashboard offers a standard range of widgets and charts, and supports the building of new dashboards through the user interface tools.

You could argue that DataTorrent RTS still needs developer resources and skills to build applications. While other vendors offer higher level development languages, and are committing to more feature-rich graphical development tools, DataTorrent sets the level of expectation appropriately and with Malhar can offer a rich library capability. The developer focus is course balanced with DataTorrent's architecture, a true dynamically scalable, fault tolerant streaming analytics platform that executes natively on Hadoop. In our view this is a significant advantage, and if your organisation has the IT skills and is actively using Hadoop, DataTorrent offers a good package for highly scalable streaming analytics.

IBM

IBM Streams is a mature product with customers across a wide range of industries including telecoms, government, financial services and automotive. The Streams offering has also expanded from on-premises deployments to cloud, with Streams now available on Bluemix, IBM's Cloud Foundry-based platform for IoT applications. Coupled with the recent announcement of Apache Quarks, an open source development platform for sensor data applications, IBM is certainly covering the bases in the streaming analytics market.

Applications on Streams are built using the Streams Processing Language (SPL), with additional support for operators built in C/C++, Java and Scala. APIs to build applications all in Java are also available. Like SQL, SPL is a high-level, declarative language, although unlike SQL, it is proprietary to IBM. That said, it would be hard to find fault with Stream's set of operators for data stream processing, which include record-by-record processing, libraries for sliding and tumbling windows, and support for a wide range of transformation, aggregation and analytics functions.

Support for building applications includes graphical development tools, with connectors for databases, NoSQL storage platforms and messaging systems. Toolkits are provided for functions such as machine learning (PMML support is included which is good to see) and geospatial operators, both packaged with the product and others available as open source libraries through github.

Performance and scalability are important weapons in IBM Stream's armoury. Benchmarks indicate impressive scalability (throughput performance and low latency) for single server and on multi-node clusters. High availability is supported, with self-healing and automatic recovery, plus the ability to ensure consistent processing with its checkpointing feature.

Other vendors may offer more out of the box in terms of visualization and dashboards for example, and put greater emphasis on product packaging for edge appliances or for wider enterprise process integration. However, there is no denying that IBM has class-leading products across its portfolio that can be integrated where required.

Going forward, it will be interesting to see how IBM's strategy for Streams and streaming analytics develops in parallel with major initiatives such as Bluemix. However, Streams is a competent and successful product that deserves its place as one of the leaders in our report.

Informatica

Informatica's vision of 'build once, deploy anywhere' includes the execution of the same analytics rules on different platforms

depending on whether the requirement is for batch processing, near real-time or real-time with streaming analytics. The Informatica Intelligent Data Platform includes streaming analytics using RulePoint coupled with Informatica's low latency messaging fabric, Ultra Messaging.

RulePoint is a real-time rules engine with stream processing capabilities rather than a pure data stream processing platform. As such, it may not exhibit the eye-watering low latency numbers claimed by some vendors. But to a large extent that's not the point. Informatica positions streaming analytics as one element of their end-to-end real-time data integration and data management ecosystem. Time to market and ease of development are key drivers. Extensive use is made of graphical platforms and templates for building and deploying rules, with productivity and reuse to the fore.

DRQL is Informatica's proprietary declarative language for specifying pattern detection rules, and where used with RulePoint, for combining rules with stream processing operations such as time windows, joins and filters. RulePoint developers use wizards with templates for building DRQL rules and to construct processing pipelines which business users can modify to cater for their changing needs. A Java SDK is available for new connectors and operators, and visualization is supported through connectors for streaming analytics to the majority of dashboard platforms.

RulePoint scales up and out over multiple nodes in the cluster, with a distributed, resilient architecture built on Ultra Messaging as the transport layer, offering low latency, guaranteed delivery and fault tolerance. Graphical tools are available for health monitoring of the operational platform. Ultra Messaging also provides the fan in, fan out mechanism for the delivery of data from change data capture (CDC) agents and Vibe Data Stream (VDS), and the connection to other Informatica platform products including PowerCenter, and with BPM where RulePoint can be integrated to trigger automated workflows. Uniquely perhaps, Informatica embeds its streaming analytics platform into its own products as pre-built business applications for real-time health monitoring, including PowerCenter, Data Quality and MDM.

In summary, RulePoint is of greatest benefit to Informatica customers seeking to extract the maximum value from their data and to add streaming analytics to their data integration and batch-based data management systems. This is where the strength of the Informatica vision plays to the maximum.

Kx Systems

Kx Systems has focused on capital markets since its launch in 1993 but is now spreading its wings into other markets that are facing similar issues of scale, data volume and performance. A subsidiary of First Derivatives Inc, an AIM-listed company, Kx has had considerable success in its core market, and can now demonstrate a rapidly growing customer base for streaming analytics in the Internet of Things, particularly in the smart energy and utilities sectors.

Kx is built on the kdb+ database, an in-memory columnar database, with the facility to optimize performance by combining the processing on in-memory data and data persisted to disk. Applications on kdb+ are built using q, a declarative, SQL-like language, with a strong mathematical library.

Q is a vector array processing language that supports the rapid, low latency analysis of time-series data, a rare facility across database platforms. Vector array processing is also highly amenable to parallelism and concurrent execution. Kdb+ offers parallel deployment across multiple partitions with a distributed capability for scale-out in a clustered environment. Resilience and automatic recovery are provided, with load balancing and replication both supported.

Strictly speaking kdb+ is not a stream processing engine. While it offers support for analysis of incoming data streams, windows-based operations are currently supported for historical data only. However it has proven credentials as a streaming analytics platform, with overall performance being more than adequate in terms of throughput and result latency.

Kx Systems' parent, First Derivatives, has built a suite of business intelligence tools on top of kdb+, including a visualization platform, and an ODBC 3 driver for connection to other popular dashboard platforms such as Tableau. Kx itself does not offer the analyst-oriented

graphical development and workflow tools offered by some vendors, but there is an active user and developer community on kxcommunity.com.

We believe that the Internet of Things will offer rich pickings for Kx with its combination of a small deployment footprint (100KB), a single platform for streaming analytics and a fast database, and the optimization of in-memory and persisted data processing. It therefore fulfills the requirements for an edge intelligence appliance, in both the data historian role, and as a streaming analytics platform for live sensor data.

Microsoft

Microsoft's Azure Stream Analytics is the only pure cloud streaming analytics platform included in this report. The majority of other streaming analytics vendors offer support for cloud deployments, often with subscription-based pricing models, but the products were not developed as true cloud-based architectures. Microsoft's is an interesting proposition, a cloud platform for streaming analytics that supports SQL queries over data streams. We feel this puts it a step above other cloud PaaS platforms for streaming analytics. Whilst Microsoft support for StreamInsight remains for on-premise deployments, Microsoft is progressing with a 'cloud first' approach and has recently launched Azure Stream Analytics as the streaming analytics function in its overall Azure feature list.

Azure Stream Analytics is built on declarative SQL with proprietary extensions for windows-based processing support. For example, Azure Stream Analytics supports GROUP BY and PARTITION BY, with support for tumbling and sliding windows. The stream processing architecture is unique – referred to as adaptive micro-batching. From the users perspective, Stream Analytics supports record-by-record processing of arriving data streams. However should the data arrival rate exceed processing capacity, the system switches automatically to micro batch mode to ensure throughput capacity is maintained at the expense of latency, while maintaining per-record timestamp processing.

Stream Analytics was designed to work in conjunction with other Azure PaaS services, with a Stream Analytics job reading from and inserting into those Azure services. This includes Power BI for visualization of streaming analytics, storage platforms such as SQL Database, Storage Blobs, Storage Tables, and Data Lake Storage, the ability to execute ML for machine learning algorithms, plus integration with Azure Service Bus Topics and Queues to drive workflow and business automation.

Cloud services are the future of the Internet of Things and it is telling that Microsoft opted for a new cloud-based architecture rather than adapt its existing mature offering for streaming analytics, StreamInsight. We believe that Microsoft is currently a step above other cloud-based solutions for data stream processing by including a SQL engine in its armoury. Of course, this is a new development and there is plenty of scope for new features and functions. And it will be interesting to see where Microsoft and other cloud vendors address the need for hybrid architectures and the Internet of Things. This is particularly important given Microsoft's strong push in the Internet of Things with Azure. Microsoft understands the importance of streaming analytics for IoT solutions based on its existing on-premise client base for StreamInsight. Therefore as a key component of Azure, we expect to see Azure Stream Analytics progress as hybrid cloud architectures evolve. If you have the in-house expertise to consider a PaaS cloud solution for streaming analytics that was built for the Cloud, then Microsoft Azure with Stream Analytics is certainly worth evaluating.

Oracle

Oracle's streaming analytics offering combines cloud and on-premise capability with a Java-based platform for delivering edge intelligence for the Internet of Things, in the Fog. Oracle Stream Analytics (previously Oracle Stream Explorer) provides a graphical JDeveloper development environment with drag and drop interfaces for the build, test and runtime deployment of real time streaming analytics applications.

Oracle Stream Analytics' focus now is providing a web tier user interface experience aimed at the business analyst, with tools to build new and enhanced applications quickly without writing code. This direction is not unique as other

vendors can demonstrate, however Oracle Stream Analytics' overall design, range of features, pattern library and contemporary look and feel, made it one of the highlights of our research. The most recent release includes a geospatial analytics library, streaming machine learning capability and a rules engine, plus an expression builder for statistical analysis.

Applications are executed in CQL (Continuous Query Language), a SQL-like language with proprietary extensions. Operators are provided for time windows, filtering, correlation, aggregation and pattern matching, with support for executing functions from the large range of high-level pattern libraries. The runtime architecture offers low latency, high throughput scale-up on Oracle hardware, or the option to deploy on a distributed cluster node by utilizing Spark Streaming. The architecture is built on what is described as an Event Processing Network (EPN) application model, with data partitioning, and high availability provided as an active-active server configurations, utilizing Coherence (in memory data grid) to maintain server state in cluster environments.

Oracle Stream Analytics offers a couple of interesting twists on the theme. First, the option to deploy over a Spark Streaming infrastructure and secondly, the use of the Coherence in-memory data grid for data stream persistence to support a resilient architecture. We're not sure support for Spark Streaming and micro-batch adds anything in stream processing terms, but we expect to see more real-time architectures that utilize in-memory data grids for both persistence and intelligent synchronization of data across enterprise platforms, and Oracle is leading the way here.

Oracle Edge Analytics is the second half of the proposition, a Java-based, small footprint platform for embedded appliance applications at the edge of Internet of Things networks. Together with Stream Analytics and support for cloud deployments, Oracle is putting together a winning combination for streaming analytics and for the Internet of Things.

SAS

SAS Event Stream Processing (ESP) is a new product introduced in 2014 that built on existing in-house knowledge of event processing architectures. ESP offers high performance streaming analytics, plus powerful inter-operability with SAS' market-leading products for advanced analytics, data quality and machine learning. SAS also had the Internet of Things market in mind as ESP was designed to deliver sophisticated streaming analytics at the network edge as an embedded appliance.

ESP supports continuous record-by-record pipeline processing of data streams over time windows. ESP's architecture delivers impressive throughput and low latency performance based on an in-memory, distributed scale-out architecture (including YARN integration) that processes events in an optimized binary format. SAS claims its 1+N failover architecture handles server and communication failures with zero data loss, and as data is not persisted for recovery, with zero impact on performance.

SAS has not gone down the SQL or SQL-like declarative language route. The recommended approach is to execute processing pipelines specified in XML through the XML Factory Server. C++ is also supported. The ESP Studio is a graphical, drag and drop environment for building and visualizing application pipelines and for deploying the applications on an executing XML Factory Server. SAS is one of the few vendors to offer explicit support for the operational test and deploy cycle, plus the ability to update a running ESP processing pipeline dynamically, without losing state or bringing down the application.

SAS is also one of the few vendors to offer real support for the execution of business rules and machine learning algorithms. PMML is supported for integration with SAS Model Manager, plus the ability to execute source code from SAS Enterprise Miner and Visual Statistics. SAS Event Streamviewer and SAS Visual Analytics can be used to visualize streaming analytics by subscribing to ESP event streams. All examples of SAS leveraging its wider product portfolio to good effect.

ESP has a growing customer base that is impressive for a new product, although of course, SAS has the advantage of a large and knowledgeable existing client base. Key use cases reported for the Internet of Things include processing at

the edge, in the gateway, as well as in the cloud or Data Center for in-vehicle analytics, manufacturing quality control, and smart energy monitoring. SAS also has customers across other markets, including real-time contextual marketing, cybersecurity, fraud, financial trading, operational predictive asset management, and risk management. SAS seems well placed to blend ESP as a maturing streaming analytics offering for cloud and edge applications, with their wider product portfolio for sophisticated analytics and machine learning.

ScaleOut Software

Many innovators in the stream processing space have focused on highly optimized architectures with pure in-memory platforms without traditional data persistence capability. This may well be the future of IT, but for now, enterprise architectures tend to combine stream processing with data persistence in some shape or form. Given the wider trend for data grids in general, and for streaming analytics platforms to be integrated with data grids, albeit for intelligent data persistence or for caching data streams and state information for resilience and recovery, we felt it useful to present the case for ScaleOut Software who is the first of the data grid vendors to tackle the streaming analytics market head on.

A data grid is a distributed, elastic scaling, self-healing and in-memory data model built on a shared-nothing architecture. IMDGs offer low latency data access and real-time query performance for big data sets, with intelligent, fine-grained control over the synchronization and persistence of grid data with underlying data stores.

Most of ScaleOut Software's customers use their in-memory data grid for distributed caching applications. However, ScaleOut has completed a number of proofs of concept with existing customers for extending the platform to streaming analytics. The extension of in-memory data grids for processing data streams can best be described as an event-driven architecture built on an in-memory, object-based representation of the world. Events are captured and stored as updates to the grid, triggering the execution of user-defined code with low latency and minimum data motion.

There are drawbacks, although we acknowledge this is early days for ScaleOut Software. There is no native ability to build stream processing operations such as time windows or time-based aggregation and transformation, nor is there a native representation of time. There is obviously application development effort required to implement these concepts.

There are however significant plus points for IMDGs. The platform offers native multi-server, distributed scale-out, with load-balancing, high availability and self-healing for resilience, and with low latency and throughput performance that is certainly competitive in the wider streaming analytics market. Analytics can be built over all data in the grid, simplifying functions such as enrichment and the joining of streaming and statics data sets.

Several vendors in this report have already demonstrated success in the streaming analytics market with the combination of a fast database with stream processing extensions. IMDGs have emerged as a key component of the enterprise architecture as IT in general pushes towards in-memory processing, real-time processes, and finer grained control over data in motion and data at rest across the Enterprise. Should ScaleOut Software persist with their strategic direction towards streaming analytics, we would expect to see native stream processing and libraries of streaming analytics operators being added to the platform. For now, we expect the offering to be most attractive to ScaleOut Software's existing customer base for their data grid, where the capability to deliver streaming analytics from the grid could be an attractive option.

Software AG

Software AG achieved the highest score overall in our evaluation and it is difficult to find too many weaknesses in their offering for streaming analytics. Built on Apama, one of the success stories from the world of CEP, and acquired by Software AG in 2013, Apama offers a comprehensive combination of performance and scalability, streaming analytics operators, graphical tools and solution accelerators for application development. Furthermore, all our reference customers for Apama highlighted stability and reliability.

That said, we believe that Software AG's primary differentiator is its ability to deliver real-time business transformation to enterprise customers. This is still an emerging market, as much about process as it is about data, where expectation is high, yet not all enterprise organizations have as yet the required levels of expertise. Software AG combines consulting services with an integrated portfolio of products for managing streaming data and real-time business processes. The suite includes streaming analytics (Apama), real-time dashboards, predictive analytics, messaging (Universal Messaging), in-memory data management (Terracotta), plus integration with business process management (webMethods).

Apama is a high performance data stream processing engine. It provides a full range of operators and window management functions for processing streaming data, with automatic query optimization, and support for both scale-up (vertically over multiple cores) and elastic scale-out (horizontally over server clusters). Apama uses a proprietary SQL-like query language called EPL (Event Processing Language). EPL includes declarative and procedural constructs with support for user extensions in Java and C++.

A long list of data feed and enterprise integration adapters are packaged as standard, plus in-built support for Terracotta's in-memory data management platform. In our view this enables more sophisticated integration patterns with databases and storage platforms, with caching for faster access, and finer grained control over database updates.

Apama Queries is Software AG's graphical drag and drop environment for application development. A visual debugger is provided, with support for stepping and breakpoints, plus a profiler for identifying performance and memory bottlenecks. REST APIs are provided for connecting to application and platform monitoring tools. Solution libraries for accelerating time to market are also available for specific target industries, including predictive maintenance and logistics.

One of the key wins in our view is the integrated engine for predictive analytics with PMML support, a feature we believe to be essential for real-time monitoring and process automation, particular for the Internet of Things. Software AG was also one of the few vendors able to demonstrate that it was getting to grips with the hierarchical architectural requirements of the Internet of Things, in particular, remote container deployments, remote application management, and giving consideration to appliance and device solutions.

Striim

Striim is a newcomer to the streaming analytics market. The company was founded in 2012 (as WebAction) but has recently rebranded following a significant funding investment in 2015. In a market where it's easy to fall short by making a grab for what appears to be low hanging fruit on every tree, Striim impressed Bloor Research with the clarity and execution of their go-to-market approach.

The founders have brought to bear their previous expertise at GoldenGate Software, establishing a beachhead built on continuous integration and ETL of data in motion into the data lake, cloud or onto Kafka, but with the advantage of streaming analytics on top. Streaming integration and analytics pipelines are built using Striim's proprietary SQL-like language (TQL) with the facility for users to add extensions written in Java. Tumbling and sliding windows are supported based on record count, time or data attributes, with support for filtering and aggregation, joining streams and historical data, pattern detection, and a small library of predictive analytics functions.

Striim provides a web-based graphical, drag and drop development environment with a contemporary look and feel. Through the user interface, a user can capture data from many sources such as databases via CDC (Change Data Capture), log files and message queues (i.e. Kafka), for example. They can then build the transformation and analytics pipelines, and connect the results to dashboards and external storage platforms. A preview capability is provided for test and debugging prior to the deployment of the application on the operational platform.

Striim may not, as yet, have the richness of stream processing capability, libraries and connectors offered by some other vendors who have been around longer, but the platform makes up for this by providing a highly scalable, elastic

architecture. The distributed execution platform combines a continuous query engine with a results cache built on Elasticsearch. Incoming data streams are sharded over the cluster for horizontal scalability, with checkpointing provided for recovery and restart from the last known good state.

Striim has made impressive progress over the past eighteen months or so since its initial product launch, leveraging its expertise in real-time data integration and CDC, and developing a horizontally scalable platform and toolset suitable for analyzing and transforming live data easily on its way to the depths of the data lake, heights of the cloud, or breadth of a Kafka distribution. Following an initial focus on financial services and telecoms, Striim, as is the trend, is now expanding into markets such as retail, manufacturing and the Internet of Things (IoT).

SQLstream

SQLstream is one of the early pioneers of stream processing and is notable for its use of ANSI SQL for executing continuous queries over arriving data streams. SQLstream has since broadened its offering, supporting application development in both Java and SQL, and has introduced StreamLab, a graphical platform (no coding required) for building streaming analytics applications over the live data streams. The SQLstream Blaze suite contains StreamLab, which is built on the core data stream processing engine, s-Server, with integrated real-time dashboards.

The query planner and optimizer is the powerhouse of their offering. SQL is declarative and therefore pliable for automatic planning and execution optimization. SQLstream has invested in automatic query optimization, and with performance features such as lock-free scheduling of query execution, has removed the need for manual tuning whilst delivering excellent throughput and latency performance on a small hardware footprint. The architecture supports distributed processing over server clusters, with redundancy and recovery options.

SQLstream scored well on performance and stream processing operators as expected, with good coverage window operations (using the SQL WINDOW clause), and support for streaming group by, partition by, union and join (inner and outer) operations for example. A geospatial analytics library is included, and an extensive library of data collection and enterprise integration connectors, including support for Hadoop, data warehouses and messaging middleware. Native support for operational data issues is included, for example handling delayed, missing or out of time order data in a way that is invisible to the user or application.

StreamLab is aligned with the wider market trend, aimed at embracing the business analyst with a set of drag-and-drop and visual tools. With StreamLab, the user interacts directly with live data streams, building pipelines of stream processing and streaming analytics operations, the results of which are pushed to integrated real-time dashboards or connected to external systems and storage platforms. Other application development aids include an Eclipse toolset for administrators and developers, and an SDK for third parties to build new connectors and data processing operators.

SQLstream is having success with its channel approach for embedding/OEMing their streaming analytics engines in cloud PaaS and SaaS platforms, and appears to be well placed to punch above its weight over the coming year.

TIBCO Software

TIBCO has extensive experience in delivering mission-critical systems built on its portfolio of products for end-to-end fast data analytics solutions. TIBCO's Event Processing suite has a strong presence in a number of verticals, including logistics, transportation, and telecommunications, with its presence in capital markets and financial services strengthened by TIBCO's acquisition of StreamBase. However, TIBCO is now packaging its offering with the Internet of Things in mind, complementing complex event processing (BusinessEvents) with predictive analytics (TERR), live datamart and dashboards (Live Datamart and LiveView), a graphical streaming analytics platform (StreamBase), in-memory data grid (ActiveSpaces), and integration with business process management. All in all, this means that TIBCO is well equipped to deliver complete IoT transformation

programs for their Enterprise customers.

At first glance there may appear to be one CEP engine too many. However, streaming analytics architectures are evolving, driven by the need to respond to the more sophisticated temporal and spatial patterns exhibited by real world IoT applications. The blending of stream processing (TIBCO StreamBase) with a rules-based CEP engine (TIBCO BusinessEvents) is a winning combination in this new world environment, allowing TIBCO to focus on deploying the strengths of both platforms in an integrated package.

The recent addition of Live DataMart is of particular interest. Live DataMart is an interactive graphical platform for business stakeholders such as business analysts to execute continuous queries over streaming data directly, and to setup alerts and visualize streaming analytics in LiveView. Live DataMart is the umbrella platform that pulls together the different capabilities in the suite into a coherent whole, including StreamBase and BusinessEvents, with predictive analytics (TERR), integration with in-memory data grid support (ActiveSpaces), plus connectors for IoT data and systems for real-time operational control applications.

TIBCO is also committed to building solution Accelerators, open source libraries delivered through the TIBCO collaboration community for improving productivity and time to deployment. There's a definite IoT bent, the first Accelerators including Connected Vehicle Analytics and Preventative Maintenance. Each Accelerator is a package of connectors, rules and streaming analytics operators that can be plugged together into applications.

In summary, enterprises looking to extract the real-time value from their fast data should consider TIBCO with its leading portfolio of horizontal and vertical solutions.

Time Compression Strategies

Ancelus is a hybrid in-memory database developed by Time Compression Strategies, a privately owned company founded in 1983. Time Compression Strategies lays claim to be the first to introduce an in-memory database (1986), the first to introduce and patent column-based processing in-memory (1991 and 1993 respectively), and were awarded the first patent for streaming analytics (1993).

Given that, and with an extensive customer base that includes many household names, you may wonder why Ancelus is not more widely known. The answer lies primarily in their focus on embedded and OEM systems for real-time analytics and control applications. Six Sigma manufacturing has been a major target market, with clients across the electronics, automotive and aerospace sectors. OEM installations in telecom network switching and cell tower call routing represent the largest installed base. Their primary applications share the common requirement of high performance, zero downtime, and reliability. The latest offering consists of the Ancelus database, plus the Real-time Route, Track and Trace (RT3) module for manufacturing and supply chain operations, and Ancelus Adaptive Analytics (A3), a library of utilities, transformation tools and statistical functions, primarily designed to support Six Sigma environments but also more general real-time applications.

Its suitability for streaming analytics lies in the unique design of the Ancelus database. Ancelus is an ACID-compliant, hybrid in-memory/disk-based database. Each data element is stored exactly once, minimizing storage requirements, and accessed through any number of logical views using either the APIs or the proprietary query language TQL (which like many languages in this report, is SQL-like rather than SQL).

At the database level, Ancelus offers time stamped data insertion into a balanced binary tree in time sorted order. Data is available to query within 250 nanoseconds regardless of table size. And as there is no concept of refresh, the current data is live and results updated with near zero latency. With the A3 streaming analytics module deployed, Ancelus can run up to a million statistical operations per second with near zero latency. A typical operation would be pattern matching within statistical limits for example.

The optimization of Ancelus for scale and real-time performance has resulted in proven credentials for streaming analytics in the most demanding of environments. However, with a focus on manufacturing and developer-level tools, it has had limited wider market adoption so far, although that is changing. New graphical development tools are available, along with a JDBC/JSON driver for standards-compliant data access. With a footprint of 100KB, dynamic platform management with zero downtime, and the combination of streaming analytics and stored data (historian) on the same platform, Ancelus has the right attributes for a strong play in the Internet of Things market.

Others

We have focused on those vendors we believe to have enterprise credentials in the streaming analytics market, and who offer a high level of performance and out of the box capability. The wider market includes PaaS cloud platforms, open source projects for data stream processing, and a growing number of companies seeking to package and commercialise open source projects.

Cloud PaaS

The leading cloud PaaS vendors include stream processing as one component of their offering: Microsoft Azure Stream Analytics, Amazon with AWS Elastic MapReduce and Kinesis, and Google's Cloud Dataflow. We opted to include Microsoft in the main report as we believe it offers greater capability out of the box for streaming analytics, in particular the use of SQL as the foundation of its stream processing language. That said, all these platforms offer a good option for organisations with the right IT skills, with access to a range of cloud-based data management components, including databases, data lakes and analytics platforms.

However, the emerging requirement for those looking towards IIoT and M2M services is the availability of edge intelligence in the Fog. There is a definite charge towards edge appliances and hybrid cloud. We expect the battle of the cloud to be one of the major developments over the coming year, between PaaS cloud platforms and those offering hybrid, hierarchical architectures.

Adding some spice to the cloud-based IoT competitive environment are some interesting variations on the theme from the world of sensor and equipment manufacturers. Most notably GE's Predix, built on Cloud Foundry and including time-series data analytics, rules and workflow. SalesForce has also entered the fray with Thunder (Heroku plus Storm, Spark and Kafka).

The impact of Open Source

Stream processing frameworks offered through the Apache Software Foundation continue to evolve within the Hadoop ecosystem. Interest in Apache Storm may be waning since Twitter announced its replacement with Heron on the grounds of performance. However, performance aside, Storm has a significant enterprise client base, is a true record-by-record stream processing framework, and is built on a topology-based paradigm for distributed processing that makes very few assumptions about the environment on which it is running.

Apache Spark has progressed to the extent that it can now make a valid claim as the leading general-purpose processing framework for Hadoop. However, the Spark Streaming API (a later addition to Spark) has yet to mature and its representation of data streams as a continuous series of RDDs is at best an approximation of stream processing. Resilient Distributed Datasets (RDD) is Spark's fundamental data abstraction and the building block for its scalable, fault-tolerant, distributed computing architecture.

Referred to as micro-batching, the RDD-based architecture is not appropriate for record-by-record processing of arriving data streams, or where data creation time and time ordering are important. It is not a suitable platform for real-time environments such as the Internet of Things, where the value and creation time of a specific record are important.

Apache Samza continues to progress without setting the world on fire, in part due to its Kafka-centric approach (Samza grew out of the Kafka ecosystem at Linkedin). However it is an older project, Apache Flink, which will be interesting to monitor. Flink, like Spark, is positioned as an alternative to MapReduce. However it can be viewed as the mirror image of Spark, that is, Flink is a stream processing framework first that offers an API for batch processing second. Flink has some interesting features, including its optimization engine and planner, declarative processing, and support for advanced stream processing features such as stream watermarks which are essential for the correct handling of gaps in data delivery.

These projects are available to download directly through the Apache Software Foundation, or through

> **We expect the battle of the cloud to be one of the major developments over the coming year, between PaaS cloud platforms and those offering hybrid, hierarchical architectures.**

enterprise vendors specialising in the packaging and commercialisation of open source offerings, for example, Cloudera (Spark Streaming), Hortonworks (Storm), Map-R Streams, Databricks (Apache Spark) and data Artisans (Flink).

An exhaustive list of Apache projects in this space is not feasible. However, of relevance to this report would be Apache Apex, DataTorrent's core processing engine, and Quarks, currently an IBM submission to Apache as a candidate for an incubator project. On a different tack, Apache Calcite is an open source framework for building expressions in relational algebra with a query planner and optimizer. That is, in common parlance, a SQL query layer, but one that can be deployed on stream processing platforms as well as static data stores. For example, Calcite has been integrated with Apache Storm and Samza to provide continuous SQL queries over data streams.

Although feature sets are improving, the open source stream processing projects remain deeply technical platforms, requiring significant development effort to deliver operational systems. They lack the graphical application development environments, product maturity and stability, reusability, richness of processing operators, and pre-built integration capability of the enterprise vendors included in this report. On the plus side, the stream processing frameworks tend to offer scale-out, distributed and fault-tolerant processing capability, plus tight interworking with Hadoop environments. Horses for courses.

Open source frameworks are free to download and use, however as streaming analytics moves into the Enterprise, it is clear that the business case is not so straightforward. Per server performance tends to be lower than enterprise alternatives, with more resources required to develop and maintain the systems. Therefore, enterprises should consider the cost of performance as a factor of hardware costs, software development effort, and on-going cost of maintenance.

“Although feature sets are improving, the open source stream processing projects remain deeply technical platforms, requiring significant development effort to deliver operational systems.”

Conclusions

Streaming analytics enables businesses to respond appropriately and in real-time to context-aware insights delivered from fast data. This is important in a world where much of the data is generated by sensors, the web, mobile devices and networks, and where the value of this data degrades rapidly over time, even over a few seconds.

Streaming analytics is forging a path through many industries, delivering real-time actionable intelligence on fraud, service quality, customer experience and real-time promotions for example. Yet even as awareness is growing as to the power of streaming analytics, and the technology required has matured, the Internet of Things is set to change the game. Streaming analytics is the core technology enabler for The Internet of Things. So far we are seeing on-premise and cloud-based platforms with real-time dashboards and alerts. Yet true disruption means business transformation, and the integration of sensor data with real-time business processes. This requires a broader solution, incorporating predictive analytics and business process management, with the ability to deliver expertise and complete transformation solutions to the enterprise.

The Industrial Internet of Things is also challenging the cloud-centric future of IT. Hybrid, hierarchical architectures are now required to deliver intelligent applications at the edge. Many of the vendors included in this report have developed platforms suitable for on-premise, cloud and appliance deployments, with several having developed specific products for edge node installations.

We have focused on vendors that we believe have demonstrated enterprise credentials with high performance, rich feature sets, and credible reference customers. We consider enterprise-level features to include comprehensive support for time-based processing, windows and streaming analytics operators, with graphical development environments, good platform performance and scalability, as well as the capability to be integrated easily with enterprise systems.

All of the vendors included in this report have significant strengths, but we appreciate that as a new market, it's not always easy for a potential user to identify which product is best for their specific use case. Streaming analytics is a broad market with significant vendor activity, including open source frameworks, vendors seeking to commercialise open source projects, and of course cloud-based PaaS offerings. We have provided as much information on each vendor and the wider market as possible. Although as always, we recommend that all users undertake their own due diligence in the shape of a proof of concept for feature functionality, integration and performance.

Finally, the future for streaming analytics is bright, but we believe a future that lies beyond the market for operational intelligence and fast analytics. Streaming analytics has to a large extent been seen as an extension of the traditional BI analytics market towards the operational coal face. This is of course true, but as IT meets real-time operations, the potential for streaming analytics is much more than merely being a conduit for perishable insights. Streaming analytics is set to become the data management platform for the real-time enterprise.

We have focused on vendors that we believe have demonstrated enterprise credentials with high performance, rich feature sets and credible reference customers.

FURTHER INFORMATION
Further information about this subject is available from
www.BloorResearch.com/update/2292

Vendor Directory

Streaming analytics is a vibrant and growing market, and as stated earlier, we have not attempted, nor would it be feasible, to analyse every vendor. In the interests of fairness and completeness, this directory offers a complete listing, with an entry for all the main vendors who offer streaming analytics products, solutions or related services.

Amazon Kinesis Streams is an AWS service for data stream ingest with usage-based pricing, plus integration with AWS services including S3, Redshift, and DynamoDB.
aws.amazon.com/kinesis/streams/

Cloudera's CDH open source Hadoop distribution includes support for Spark Streaming.
www.cloudera.com

Cisco Streaming Analytics (CSA) is a core component of Cisco's Fog Node strategy for edge analytics, available standalone and as an embedded appliance within Fog Nodes.
www.cisco.com

data Artisans leads the development of Apache Flink, an emerging open source framework for distributed stream and batch processing.
www.data-artisans.com

Databricks was founded by the creators of Apache Spark, to monetize Spark and Spark Streaming as a commercial venture.
www.databricks.com

DataTorrent RTS is a distributed streaming analytics platform on Hadoop. DataTorrent recently submitted its RTS engine as an open source Apache project called Apex.
www.datatorrent.com

EsperTech, Esper is one of the more mature event stream processing platforms, offering an enterprise-level product built on its open source complex event processing engine.
www.espertech.com

Google's Cloud Dataflow is a managed service for stream processing in Google's Cloud Platform, integrated with other cloud portfolio products such as BigQuery.
cloud.google.com/dataflow

Guavus targets telecoms and IIoT with Reflex, a platform for streaming analytics solutions built on Hadoop, Spark and Yarn.
www.guavus.com

The **Hortonworks** Data Platform includes Apache Storm as the stream processing component of its Hadoop distribution.
www.hortonworks.com

IBM Streams is available on-premise and on Bluemix, IBM's hybrid cloud platform. Streams is a mature product with an extensive customer base.
www.ibm.com

Informatica's Intelligent Data Platform has RulePoint as its streaming analytics engine integrated with Ultra Messaging, plus a 'build once, deploy anywhere' capability for executing analytics across its portfolio.
www.informatica.com

Impetus Technologies' StreamAnalytix is a `pluggable architecture for building streaming analytics solutions on open source frameworks such as Spark Streaming, Storm and Esper.
www.impetus.com

Kx Systems, a subsidiary of First Derivatives Inc, has an extensive client base built on kdb+, a fast in-memory columnar database with streaming analytics capability.
www.kx.com

MapR Streams is a framework for incorporating platforms such as Spark, Storm, Apex and Flink, delivered as part of MapR's Converged Data Platform.
www.mapr.com

Microsoft offers Azure Stream Analytics as a pure cloud-based streaming analytics platform, plus continued support for its on-premise offering, StreamInsight
www.microsoft.com

OneMarketData's OneTick offers integrated products for streaming and historical data analytics, primarily for Capital Markets.
www.onetick.com

Oracle Stream Analytics is available both on-premise and in the cloud, complemented with a Java-based platform for IoT edge appliances.
www.oracle.com

SAP packages its event stream processing platform (ESP) as a standalone product and as an add-on to SAP HANA as Smart Data Streaming (SDS).
www.sap.com

SAS Event Stream Processing (ESP) offers inter-operability with SAS' market-leading products for advanced analytics, data quality and machine learning.
www.sas.com

ScaleOut Software offers event stream processing capabilities as part of its operational intelligence platform that adds a data-parallel compute engine to its core in-memory data grid.
www.scaleoutsoftware.com

Software AG integrates Apama, its mature streaming analytics platform, with its in-memory data grid, predictive analytics and fast messaging platforms, offering complete business transformation solutions for real-time systems.
www.softwareag.com

SQLstream Blaze is built on a SQL-compliant data stream processing engine with StreamLab, a visual, drag and drop platform for building applications from live data streams.
www.sqlstream.com

Striim is a relative newcomer, building on its data integration credentials to offer a scalable streaming analytics platform with a user-friendly graphical interface.
www.striim.com

TIBCO offers both StreamBase and a rules-based CEP engine, BusinessEvents, as core components of its end-to-end portfolio for fast data analytics.
www.tibco.com

Time Compression Strategies, experts in industrial automation and IIoT, offers Ancelus, a fast in-memory database with an innovative architecture for streaming and historical data analytics.
www.tcsdb.com

Vitria focuses its streaming analytics capability on the operational and process intelligence markets.
www.vitria.com

WSO2 offers complex event processing as an extension of its open source middleware platform.
www.ws02.com

About the author

RONNIE BEGGS
Senior Analyst

Ronnie Beggs specialises in the field of real-time IT solutions and how industry is combining internet, cloud computing, big data and stream processing technology to deliver innovative applications. With a background in embedded systems as well as enterprise IT technology, Ronnie has a particular interest in how data management platforms are replacing both hardware and legacy real-time systems in industries such as telecommunications and the Internet of Things.

Ronnie is relatively new to the analyst community, bringing more than twenty years experience of product management, marketing and business development in the real-time software business. Ronnie's career spans a range of industries, from safety critical software in the aerospace industry, operational performance and data management in telecommunications, and real-time monitoring and control applications for the Internet of Things. Ronnie has worked for a number of successful start-ups, from early stage through to acquisition.

Outside of work, Ronnie enjoys cycling, rugby (definitely watching rather than playing these days), going to the theatre as often as possible, throwing pots on the wheel, and being endlessly amused at how the media and politicians misunderstand and misrepresent technology, data and statistics

Bloor overview

Bloor Research is one of Europe's leading IT research, analysis and consultancy organisations, and in 2014 celebrated its 25th anniversary. We explain how to bring greater Agility to corporate IT systems through the effective governance, management and leverage of Information. We have built a reputation for 'telling the right story' with independent, intelligent, well-articulated communications content and publications on all aspects of the ICT industry. We believe the objective of telling the right story is to:

- Describe the technology in context to its business value and the other systems and processes it interacts with.
- Understand how new and innovative technologies fit in with existing ICT investments.
- Look at the whole market and explain all the solutions available and how they can be more effectively evaluated.
- Filter 'noise' and make it easier to find the additional information or news that supports both investment and implementation.
- Ensure all our content is available through the most appropriate channel.

Founded in 1989, we have spent 25 years distributing research and analysis to IT user and vendor organisations throughout the world via online subscriptions, tailored research services, events and consultancy projects. We are committed to turning our knowledge into business value for you.

Copyright and disclaimer

This document is copyright **© 2016 Bloor**. No part of this publication may be reproduced by any method whatsoever without the prior consent of Bloor Research. Due to the nature of this material, numerous hardware and software products have been mentioned by name. In the majority, if not all, of the cases, these product names are claimed as trademarks by the companies that manufacture the products. It is not Bloor Research's intent to claim these names or trademarks as our own. Likewise, company logos, graphics or screen shots have been reproduced with the consent of the owner and are subject to that owner's copyright.

Whilst every care has been taken in the preparation of this document to ensure that the information is correct, the publishers cannot accept responsibility for any errors or omissions.

Bloor Research International Ltd
20–22 Wenlock Road
LONDON N1 7GU
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloor.eu**
email: **info@Bloor.eu**