SOLUTION BRIEF

Data Center
Advanced Analytics

Streaming Analytics in the Real-time Organization

How real-time insights unlock new business value

Table of Contents

"We live in a real-time world, and so too should enterprise applications"

– **Mike Gualtieri**, VP & Principal Analyst Forrester Research[1].

The Broadening Opportunity for Streaming Analytics

The number of enterprise opportunities for streaming analytics is growing rapidly, as businesses across verticals look to process and correlate data from multiple sources to obtain insights in real-time. Streaming analytics is clearly positioned to move into the mainstream as more enterprises capitalize on in-memory databases and streaming analytics to process, correlate and analyze data in real-time – whether at the edge, on premise or in the cloud.

The consequence is a broader, and growing business demand for the insights streaming analytics can provide, as decision makers at all levels look to switch from backward- to forward-looking decisions. Many big data solutions that require real-time insights today are being driven out of line of business (LOB) divisions, not just the CIO's office, as organizations recognize that speed of insight is a business differentiator, enabling nimbleness and competitiveness.

In part, this is being driven by the Internet of Things. The proliferation of 'smart' devices and sensors, each of which is capable of generating a great deal of new data, directly drives the need for streaming analytics. According to analyst firm Gartner*, 20 billion internet-connected things would be in use worldwide in 2020, and more than 65 percent of enterprises (up from 30 percent in 2017) will have adopted IoT products.[2]

In this paper we consider the growing role of streaming analytics, illustrated by experience of key players in the space – SAP* and SAS*. For instance, SAS Event Stream Processing (ESP) has been used in a wide variety of scenarios including predictive analytics. Here we look at how ESP can impact operational processes. Meanwhile SAP HANA* can work in tandem with in-memory processes to drive insights across the organization, right up to board level. We also look at the potential of Apache Spark*, which is expected to grow at a compound annual growth rate (CAGR) of 67 percent between 2015 and 2021[3]. A disruptive technology, its role in connecting streaming analytics with machine learning and artificial intelligence (AI) is critical.

Enterprises have several options to pivot to a "real-time insights enterprise", and Intel has been at the forefront in optimizing streaming analytics workloads. In this paper we review the opportunity and give you direction in where to start looking for your next steps to streaming analytics.

Why is Streaming Analytics Such a Hot Topic?

Traditionally, data management and analytics capabilities have been largely batch-oriented, working with limited sets of historical data. Greater data collation, faster throughput and reduced analytics latency enable insights to be generated and delivered in seconds, opening the door to new kinds of decision making, new customer experiences, and greater automation of repetitive tasks due to several, converging factors:

- **Connectivity** has exploded in terms of both bandwidth and reach, enabling access to a widening range of data sources and device types.
- **Edge devices** from Internet of Things (IoT) sensors to local processing, have become smarter and more powerful, enabling smarter data collection, collation and delivery. According to Gartner, 20.4 billion connected things will be in use by 2020[2].
- **Core computing** has dramatically increased in power, both in data centers and in the cloud, creating massively scalable processing architectures.
- **In-memory processing** is leveraging faster and denser capabilities such as Intel® Optane™ technology, delivering a revolutionary leap forward in decreasing latency, accelerating applications for fast caching and storage, reducing transaction costs for latency-sensitive workloads and increasing scale per server.
- **Analytical software tools** have adapted and evolved, together with line of business applications, to provide insights at the point of need. For example, big data analytics platforms including Apache Spark* use machine learning techniques to process, correlate and analyze streaming data from multiple sources in real-time.

Streaming Analytics – From Board Room to Front Line

As streaming analytics becomes applicable to an increasingly wide range of use cases, the benefits of harnessing its insights can be felt across the organization, from the board room to the front line. The following scenarios illustrate the power of streaming analytics solutions.

> "Machine intelligence is... a prediction technology, so the economic shift will center around a drop in the cost of prediction... [with] two other well-established economic implications. First, we will start using prediction to perform tasks where we previously didn't. Second, the value of other things that complement prediction will rise."[4]

– Professor Ajay Agrawal
(Economist at the University of Toronto) and colleagues

Operational Improvements through SAS Event Stream Processing

Software company SAS has been in the big data and analytics business for over 40 years. With its advanced ESP engine, SAS Analytics for IoT* can process millions of events per second, driving the potential for new operational models through:

- Identifying and assessing problems that would otherwise go unnoticed
- Forecasting impending failures and generating real-time notifications
- Filtering, aggregating, and correlating data to delineate patterns of interest

These capabilities mean not only that issues can be addressed proactively, but they also enable new working practices. For example, air filters in corporate buildings are typically changed every six months based on average life expectancy, without taking into account contextual factors (such as pollen count). Use of sensor-based data and ESP can lead to both better air quality and less waste, through use of proactive maintenance plans rather than arbitrary schedules.

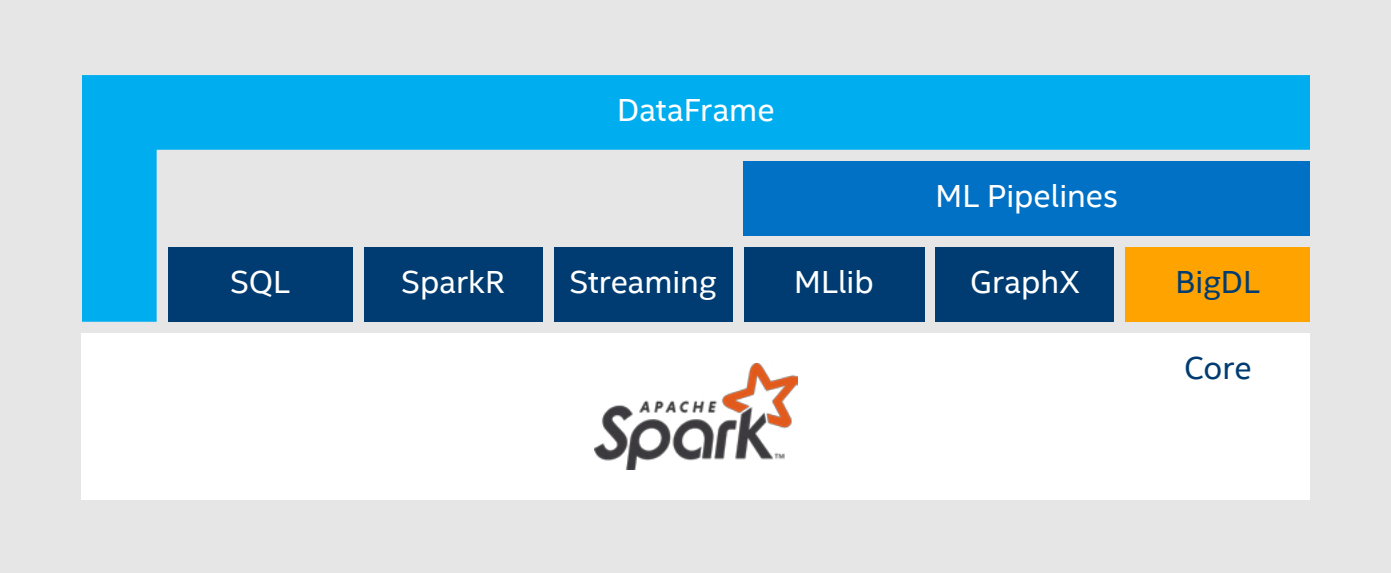

Figure 1. BigDL in the Apache Spark* stack

Board-level Insight through SAP HANA and In-memory Databases

Intel and SAP have been working together since 2005 to deliver better performance for SAP applications running on Intel® architecture. The SAP HANA platform takes full advantage by combining columnar data storage, massively parallel processing (MPP), and in-memory computing.

SAP's Digital Boardroom application delivers real-time insights from across the business to senior executives, driving decision making across manufacturing and logistics, HR and resourcing, finance and sales to ensure that stakeholder value is maximized. This example illustrates the opportunity for organizations to benefit from the use of streaming analytics from the top down as digital transformation creates new sources of data across enterprise.

As streaming analytics moves from an exception to a core component of the data architecture and infrastructure, businesses can adopt a 360-degree view not only of their customers but also of themselves. Gartner describes this as a corporate 'digital twin' vision[5], driven as much by pain as by gain: while 'good' data can add value[6], according to one estimate, in 2016 bad data cost the United States $3.1 trillion[7].

Fraud Detection via Deep Learning using Apache Spark*

Apache Spark has emerged as a game-changer for big data processing in recent years. With its open-source, in-memory architecture, Spark can process and analyze not only batch data but massive volumes of streaming data in real-time: its high performance comes from the ability to do in-memory processing instead of saving data to hard disks. According to research firm Wikibon, the adoption of Apache Spark is gaining momentum and will reach a growth rate of 72 percent by 2019[8].

A key benefit of Spark is its ability to handle an entire analytics pipeline, as well as train machine and deep learning models at scale. To make this workflow simpler and more efficient for deep learning deployments, Intel is leading an open source initiative known as BigDL, a distributed library for building deep learning applications with Apache Spark.

Using BigDL, Apache Spark enables streaming analytics to feed training and inference algorithms. As shown in Figure 1, unlike a number of other libraries for deep learning, BigDL is native to Apache Spark: using BigDL, you can write deep learning applications as standard Spark programs that run on existing Spark or Hadoop* clusters. BigDL therefore accelerates time to value, reduces total cost of ownership (TCO), and improves ease of use[8].

In financial services for example, an Apache Spark and BigDL-based model powered by streaming analytics enables expansion beyond electronic trading, enabling deeper insights for risk management, regulatory compliance and fraud detection. If fraud can be identified in real-time, it can be prevented as it happens, resulting in greatly reduced costs and much higher levels of protection.

Streaming Analytics Use Cases

As the potential for streaming analytics increases, different industries are adopting it in new ways, delivering unprecedented value to stakeholders as the following examples show.

Clinical Healthcare

Real-time analytics provides the ability to monitor patient safety, personalize patient results, assess clinical risk and to reduce patient readmission—improving organizational efficiency and the patient experience. For example, remote monitoring and rapid response to clinical events can significantly improve quality of care for chronically ill patients.[9]

Transportation

Logistics companies can use streaming data from sensors to pre-empt engine failures, reducing the risk (and considerable expense) of making repairs on the road[10]. Meanwhile data from satellite and roadside systems can be monitored in real time and fed into a centralized dashboard; or returned to vehicles to drive advanced driver assistance systems.[11]

Retail Inventory and Behavioral Monitoring

Streaming analytics can improve real-time inventory tracking and understanding of how customers act in-store or online, driving operational efficiency, higher sales volumes, better security and enhanced customer satisfaction. It can also identify anomalies in context-based data (such as market trends) that have an impact on business performance.[12]

Call Center Analytics and Customer Churn

Using call center logs, data from sensors, online interactions, interactive voice response, and IT support systems, enterprises can model customer behaviors, churn prediction and other insights[13]. Predictive models help increase staff retention and reduce considerable recruitment overheads.

Network Traffic Monitoring and Fraud Detection

Cybersecurity is among the most critical threats enterprises face today, as increasingly sophisticated hackers continuously seek vulnerabilities they can use to steal data. Streaming analytics with its machine learning capabilities can help detect anomalies and provide early warning of attacks.

Financial Fraud Detection and Credit Transaction Monitoring

Streaming analytics can aggregate transactional data to improve business health and reveal insights about customer behaviors. Machine learning models can help to predict fraudulent transactions, enabling faster reaction times and reducing exposure to risk.[14]

More Flexible Insurance Models

Real-time transmission data opens the door to new capabilities and business models for insurance companies, for example 'pay as you drive' insurance. Fitness data can also be used as input to health insurance models and product innovation, for example 'rewarding' customers for undertaking activities.[15]

Smart Operations and Supply Chain

Analytics-enabled predictive maintenance enables servicing with minimal disruption and cost, optimizing supply chains and forward capacity planning for spare parts and raw materials.[16] Streaming analytics can speed up production and maintenance cycles, increase quality and drive incremental improvement programs.[17]

Home and Business Automation

Smart meters can manage energy efficiency based on occupancy, with savings up to a third in some cases[18], and reduce equipment risk in homes and businesses, for example pre-empting refrigeration failure in a convenience store. They can also feed data to secondary substations, ensuring energy is available to meet demand.[19]

IT and Business Operations

Systems data analytics enables better IT responsiveness and capacity planning, and can feed directly into continuous delivery and automated management processes. Streaming analytics from devices enables detection and prediction of failures, driving IT efficiency across infrastructure, operations and resources, and into business processes.

First Steps to Streaming Analytics

As we have seen, a wide range of opportunities exists for streaming analytics. So, as well as having a broader architectural vision, how should you go about developing a strategy that can deliver on the vision of streaming analytics? Based on the experience we have gained working with customers at various stages on their streaming analytics journeys, we have distilled the following guidance.

1 Design security and quality into connectivity

A first step is to define how you are authorizing and connecting to devices, assuring that information is being correctly and securely reported, that devices are kept up-to-date and that information fits within governance and compliance policies. These aspects are directly in the domain of IT as it is unlikely the business unit will think of it first. Such questions are equally about future safety, as endpoint devices and sensors are difficult to change once in place.

Security is a number one challenge, as anything that is connected to your network is a potential avenue for attack. How are you going to deploy and maintain the devices in a secure manner? How are devices going to identify themselves and demonstrate they haven't been compromised? Sensor data needs to be secure, encrypted and only visible to the right people at the right time.

You also need to consider how you will assure data quality over time. For example, a weather reporting company may be running many thousands of sensors, but they will only be able to report consistently if their environment stays the same. However, walls can be erected, or sensors can be moved without notice or knowledge. If the environment around the sensor changes, it will affect the data being produced, compromising potential insights. With Wind River technologies*, Intel can help secure data at the edge, and between the gateway and the cloud or data center.

2 Consider a hybrid cloud strategy

Cloud-based services are particularly useful for organizations looking to test their streaming analytics data architecture. While you can design a full-scale solution, you can implement it on a smaller scale in the cloud, verify its operation and see how it is used, what challenges it presents, and what areas need enhancement. Then you can redefine the architecture and slowly expand the usage within the cloud as far as possible. Note that some industries (including finance) emphasize data sensitivity, or have regulatory needs that make use of cloud-based services more difficult.

Many organizations are looking to a hybrid strategy, in which they need to decide what is stored and processed in the cloud versus locally. Some organizations are looking to cloud to be their data center of the future, building the whole new data architecture as well as the analytical infrastructures in the cloud from the start. Without a data center to manage, they can focus on business problems. And smaller companies can also benefit from analytics-as-a-service offerings, which are much faster to deploy and have lower total cost of ownership (TCO).

3 Think big, start realistic

A significant risk can be created if organizations try to do something too big without prior experience. To counter this, pick a clearly defined goal with a clear return on investment: this helps start establishing the value of streaming analytics within the company, build credibility and gain expertise. For example, in a manufacturing environment, rather than trying to instrument the entire factory, it is worthwhile to focus on which equipment is most prone to failure.

The same challenge arises for connectivity, as the IT network needs to be extended effectively and securely to a broad range of locations that typically haven't had computational devices in place. It is worth testing a model with a small number of devices, sensors or gateways before scaling up. An additional area to be taken into account involves operational management overhead: it is much easier to establish a need by testing it out on a small scale, than find oneself poorly resourced after a major rollout.

4 Decide, and review, what data to stream

Given the volumes of data in today's world, it is not possible to transmit or process everything. While you don't need to keep derived data, managing potential bottlenecks needs to be balanced with the fact that data can drive new, sometimes unexpected insights. In DNA sequencing for example, data initially seen as unnecessary was found later to have a purpose, meaning analysis needed to be re-done. As machine and deep learning techniques identify new relationships between data items and produce new and improved models, information seen as irrelevant can become very important.

On an ongoing basis, test models and data to test assumptions are correct, and to confirm confidence levels in results, particularly if data is likely to change. New models can be pushed to edge devices to deliver better results, creating and producing better and more accurate insights. Meanwhile, it is important to capture as much data as is economically feasible at any moment in time. This needs to be balanced against policy: for example, in banking, what data will you need to show in case of a lawsuit, or to demonstrate compliance?

5 Plan to change the business

Streaming analytics is not just about gathering, storing and analyzing data, but also operationalizing the results. It is one thing to make insights available 'to the masses' but they then need to act on it – and will not necessarily be geared

up to do so. For example, a maintenance department may be structured around schedules, rather than reactive or predictive maintenance.

If you're not changing what you're doing based on the data that you're getting, the rest of the steps are useless. So, how are you taking those results and feeding them back to the people who will actually adjust how the company is working, how they're interacting with the customers or doing maintenance, to drive real-world savings? This may change the make-up of the work force.

As you integrate streaming analytics into your enterprise, it is important to work closely with specialists who understand the inception of new technology and have the know-how to align it to your own, staged transformation approach. Help is at hand from organizations including Intel to assure you deliver in your shorter- and longer-term ROI goals.

What Intel Brings to Streaming Analytics

Streaming analytics demands a future-ready platform that can support complex data management, high performance, low latency, and the ability to operate at scale. As we have seen from the examples, highly efficient hardware combines with new, efficient data analytics software to broaden the range of opportunities for streaming analytics into entirely new areas. From a hardware perspective, Intel continues to operate at the forefront of advances in streaming analytics:

- Intel® Xeon® Scalable processors achieve exceptional resource utilization and agility, with clock speeds and Intel® technologies to deliver amazing analytics performance without relying on high-core-count processors.
- Intel® Optane® technology is a unique combination of Intel® 3D XPoint™ Memory Media, Intel Memory and Storage Controllers, Intel Interconnect IP and Intel® software. Together these building blocks deliver decreased latency and accelerated systems for workloads demanding large capacity and fast storage.
- Intel® Memory Drive Technology is a software-defined memory (SDM) technology that expands system memory transparently. Combined with an Intel® Optane™ Solid State Drive (SSD), it integrates the SSD into the memory subsystem and makes it appear like DRAM, increasing memory capacity beyond DRAM limitations.

At Intel, we continue to optimize across hardware, storage and software to handle massive volumes of streaming data, from ingest through to processing and delivering both performance and scalability up to terabytes of data. Watch this space as we continue to make new announcements on how to use Intel technologies to deliver on your streaming analytics goals.

To further accelerate the process of selecting and deploying the hardware and software needed for today's broad array of analytics scenarios, Intel® Select Solutions offers a variety of flexible, easy-to-deploy, scalable solutions optimized for performance.

With Streaming Analytics, Prepare your Organization for Real-time Insights

The modern data explosion, combined with an increasing need for decisions based on real-time data, makes real-time streaming analytics bolstered by machine learning an absolute necessity for firms who want to remain competitive. The overall goal is decision-making confidence – as being able to predict more accurately drives new business models and opportunities, increasing competitiveness and enabling a winning edge. Explore the growing library of Intel® Select Solutions.

Starting down the track of streaming analytics is perhaps one of the most important business decisions your organization could make. We would recommend the following further reading as you continue to develop your understanding of what streaming analytics can bring:

Solution Brief: Business Intelligence at the Edge – End-to-end IoT Portfolio from Intel and SAP

Solution Brief: Intelligent Analytics for Smart Machines

Product Brief: SAP HANA* Real-Time Analytics Software and Intel® Technologies

Solution Brief: Accelerate Time to Insights and Enable the Ultimate Shopping Experience

Intel® AI Academy: BigDL: Distributed Deep Learning on Apache Spark*

Explore Intel advanced analytics solutions at **intel.com/analytics**

Solutions Provided By:

[1] The Forrester Wave™: Streaming Analytics, Q3 2017 https://www.forrester.com/report/The+Forrester+Wave+Streaming+Analytics+Q3+2017/-/E-RES136545

[2] Leading the IoT - Gartner Insights on How to Lead in a Connected World, edited by Mark Hung, Gartner Research,2017, https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf

[3] Market Analysis Report, Apache Spark Market Forecast 2017-2020, May 2017, https://www.marketanalysis.com/?p=159

[4] The Simple Economics of Machine Intelligence, Ajay Agrawal, Joshua Gans, Avi Goldfarb https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence

[5] Prepare for the Impact of Digital Twins, Gartner, https://www.gartner.com/smarterwithgartner/prepare-for-the-impact-of-digital-twins/

[6] How Real-Time Data Boosts the Bottom Line. http://cioinsight.com/it-management/innovation/slideshows/how-real-time-data-boosts-the-bottom-line.html

[7] The Cost of Bad Data. http://information-age.com/cost-bad-data-123465972

[8] Accelerate Deep Learning Workloads on Apache Spark with BigDL, Intel, https://itpeernetwork.intel.com/accelerate-deep-learning-apache-spark-bigdl-2/

[9] Maximize Data Value with a Real-Time Streaming Analytics Solution, Intel, https://www.intel.com/content/dam/www/public/us/en/documents/solution-briefs/real-time-streaming-analytics-brief.pdf

[10] Driving IoT Enabled Predictive Maintenance, Cloudera, https://vision.cloudera.com/driving-iot-enabled-predictive-maintenance/

[11] Traffic Data Monitoring Using IoT, Kafka and Spark Streaming, https://www.infoq.com/articles/traffic-data-monitoring-iot-kafka-and-spark-streaming

[12] Ebay's Analytics and Business Intelligence, https://www.intel.com/content/www/us/en/processors/xeon/business-intelligence-analytics-ebay-signals-video

[13] Bloor Market Report – Streaming Analytics 2016, https://kx.com/media/2016/12/Bloor-Market-Report-on-Streaming-Analytics.pdf

[14] Capitalizing on Big Data Analytics for the Insurance Industry, StackIQ, https://cdn2.hubspot.net/hub/173001/file-18488782-pdf/docs/stackiq_insuranceind_wpp_f.pdf

[15] UnitedHealthcare and Fitbit to pay users up to $1,500 to use devices, Fitbit co-founder says, https://www.cnbc.com/2017/01/05/unitedhealthcare-and-fitbit-to-pay-users-up-to-1500-to-use-devices.html

[16] Rapid, Actionable Analytics with Intel and Software AG, https://www.intel.com/content/www/us/en/processors/xeon/softwareag-actionable-analytics-video.html

[17] IT@Intel: Using Big Data in Manufacturing at Intel's Smart Factories, https://www.intel.com/content/www/us/en/it-management/intel-it-best-practices/using-big-data-in-manufacturing-at-intels-smart-factories-paper.html

[18] Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings, Junjing Yang, National University of Singapore https://www.researchgate.net/publication/286455523_Review_of_occupancy_sensing_systems_and_occupancy_modeling_methodologies_for_the_application_in_institutional_buildings

[19] Digitizing power utilities – Business transformation driven by advanced analytics, Intel, https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/iot-digitizing-power-utilities-paper.pdf

0718/JS/CAT/PDF 337754-001EN