Tap the Value Hidden in Streaming Data: the Power of Apama Streaming Analytics* on the Intel® Xeon® Processor E7 v3 Family

To identify and seize potential opportunities, in addition to recognizing and neutralizing threats, enterprises need a real-time analytics platform that can turn relentless streams of data into meaningful and actionable insights. When captured in the right context, data streams can help identify events that enterprises can analyze to tap hidden value from numerous sources such as financial markets, mobile devices, Internet of Things (IoT) sensors, clickstreams, and business transactions.

This white paper discusses the business value of real-time analysis of streaming data using the Software AG Apama Streaming Analytics* platform powered by the Intel® Xeon® processor E7 v3 family. It illustrates a potential use case for streaming analytics in retail, and it presents test results that show 59-percent higher throughput with Apama running on a server powered by the Intel Xeon processor E7 v3 family compared to when running on a previous-generation processor.[1] A second set of tests, where events were sent to the Apama Correloator over a network, demonstrate 2.64 times faster throughput for an Apama server powered by the Intel Xeon processor E7 v3 family and a 40 gigabit Ethernet (GbE) Intel® Ethernet Controller, compared to a server powered by a previous-generation processor and a 10GbE controller.[2]

The Value of Real-time Analysis of Streaming Data

Streaming data conceals important events and trends that represent business opportunities and potential liabilities such as customer shopping patterns, customers' locations, credit-card fraud, and financial-market trading abuse. More and more enterprises are employing event processing systems that use streaming data analytics to uncover these events and trends. These enterprises are able to quickly and effectively respond to constantly changing business conditions. Use cases include:

- **Customer-experience management:** Detect and take real-time action on customer behavior.
- **Real-time personalization:** Provide sales associates with relevant insights to connect with customers.
- **Fraud detection:** Reveal fraud and identify unusual behaviors and locations.
- **Foreign exchange e-commerce:** View connected markets in a single screen and offer the best prices while managing risk.
- **Market surveillance and monitoring:** Rapidly detect abusive or faulty trading and respond quickly.
- **IoT:** Harness the vast streams of sensor data and visualize data in motion.

To extract value from enormous and exponentially growing data streams that get stale quickly, enterprises need

Table of Contents

real-time analytics on dynamic data and technologies that support performance at scale. The task might seem daunting, but the rewards for business could be huge. McKinsey & Company estimates that by the year 2025 IoT applications could yield as much as $2.3 trillion in new economic value worldwide—and that's in manufacturing alone.[3]

The Software AG Apama Streaming Analytics platform running on servers powered by the Intel Xeon processor E7 v3 family can quickly turn never-ending data streams into meaningful real-time metrics. This solution enables intelligent, automated actions and visualizations to support non-automated decisions.

"IoT could generate $2.3 trillion in new value by 2025 [3]."

The Software AG Apama Streaming Analytics platform running on servers powered by the Intel Xeon processor E7 v3 family can quickly turn never-ending data streams into meaningful real-time metrics. This solution enables intelligent, automated actions and visualizations to support non-automated decisions.

Apama: A Comprehensive Event Processing Platform

The event-driven design of Apama Streaming Analytics platform combines event processing, messaging, in-memory data management, and visualization in a single comprehensive platform that enables enterprises to capture streaming data from any device, perform analytics on the data, and identify meaningful events and patterns. The in-memory architecture of the Apama

Figure 1. Components that comprise the Apama Streaming Analytics* event-processing architecture

platform enables real-time processing of enormous data volumes that are moving at extremely fast speeds. The Apama platform includes a development environment and integrates connection to external systems.

"It is not surprising that Software AG has the highest 'current offering' score [in the big data streaming analytics market]" – Forrester Research, Inc.[4]

The Apama platform has three major components shown in Figure 1:

- **Apama tools:** A rich, intuitive development environment (including real-time graphical dashboards for monitoring application execution) enables business analysts and IT developers to quickly and easily build, test, and evolve Apama applications.
- **Apama Correlator*:** The run-time engine that consumes all data sources, executes the required applications, combs the data for event patterns, and delivers insights and automated actions in real-time.
- **Apama integration framework:** Integrates Apama with external systems such as event messaging environments, databases, and existing application environments that can provide real-time feeds and static data to Apama and receive and visualize data from Apama.

The Software AG Apama Correlator*

The execution engine for the Apama platform is the Apama Correlator, where Apama applications are inserted and removed dynamically at run time without disrupting other running applications. The Apama Correlator consists of multiple parallel-execution containers that are called "contexts" as shown in Figure 2.

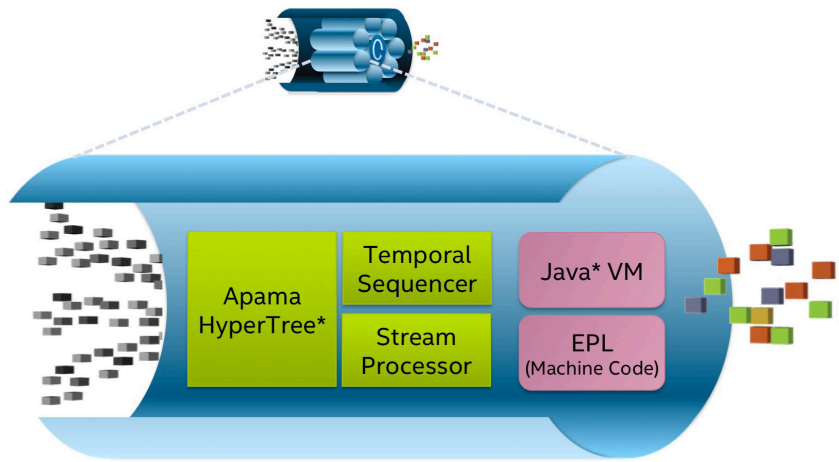

Figure 2. A number of contexts within a single Apama Correlator* with the details of a single context

Apama applications are modular compositions that break down event monitoring, analysis, and action into independent but logically related segments. This modular design enables thousands of contexts to operate simultaneously within the Apama Correlator, and each context can process any number of Apama applications. The Apama Correlator has a highly optimized thread scheduler that controls the execution of the numerous contexts by mapping the contexts to a smaller number of processor threads. This allows Apama to easily, elastically, and efficiently scale over all the available processors and cores within a single host. The Apama Correlator has several components shown in Figure 2.

Apama HyperTree*: Index the Query, Not the Data

A key element of the Apama Correlator is Apama HyperTree*, an innovative indexing technology that indexes the query rather than the data. This approach overcomes significant performance overhead and latency of the traditional indexing approach and is ideally suited to event processing of big data. Apama HyperTree contains data structures and algorithms designed for high performance multi-dimensional event filtering. Unlike traditional systems, Apama HyperTree allows indexing of many fields within an event, which enables high-speed matches. The detection of events is not dependent on index recalculation as would be required to perform a similar operation with a database. Event patterns can be detected instantly without additional processing. The Apama HyperTree indexing technology enables enormous scalability with minimal performance degradation—an Apama benchmark test shows that a 1,000-percent increase in the number of event patterns being monitored causes only a 15-percent reduction in throughput.[5]

Temporal Sequencer

The Apama Correlator includes a temporal sequencer that performs temporal correlations. To understand how this works, consider an application that seeks occurrences of event A followed within 500 milliseconds by event B. In this case the Apama Correlator initially only looks for event A. Only when event A is detected does the Apama Correlator begin to look for event B and only for a maximum of 500 milliseconds as specified in the scenario. If no event B is detected in that time, the application terminates.

Stream Processor

The Apama Correlator also contains a stream processor that stores and organizes events that occur within a time interval. The stream processor orchestrates the execution of real-time analytics over these event intervals and provides a powerful, efficient, and easy-to-use set of capabilities for defining real-time analytics and for identifying relational-style patterns across multiple streams of data.

Application Processing

Once an event pattern is identified, the Apama Correlator passes the query results to the application that requested that requested the query. Applications written in the Apama Event Processing Language (EPL) are executed as native machine code.6 Applications written in Java* are executed within the Java Virtual Machine (JVM*) that is built into the Apama Correlator. Apama benchmark tests show that EPL applications perform up to 25 percent faster than Java and C/C++ applications.[5]

Apama Integration Framework

To facilitate the connection of event data sources with the Apama Correlator, the Apama platform includes an integration framework called the Integration Adapter Framework (IAF). The IAF enables easy creation of software adapters to interface the Apama Correlator with middleware buses and other message sources. IAF adapters are protocol-neutral and can communicate with third-party messaging systems, extract and decode self-describing or schema-formatted messages, and transform them into Apama events. All events, regardless of their sources, are converted to the Apama format, which allows Apama to natively support correlations that span different external data formats. More importantly, IAF adapters are bi-directional—they can also convert Apama events into representations required by third-party messaging systems such as an order to a trading exchange.

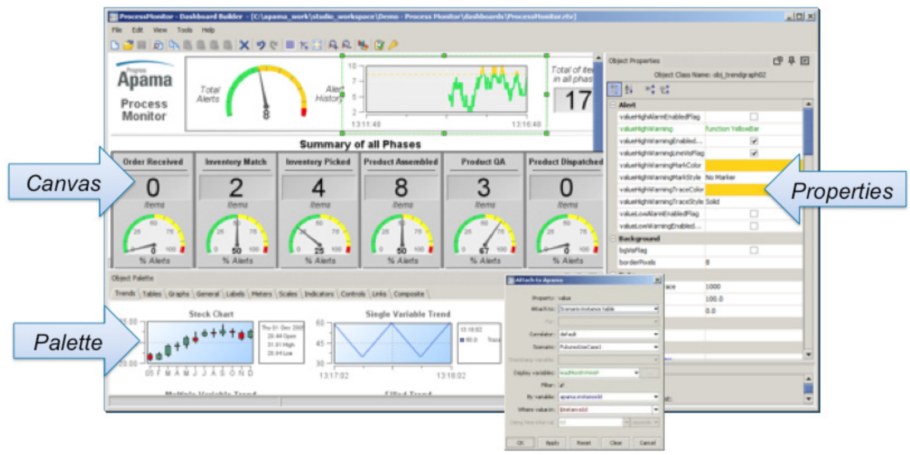

Figure 3. Apama Dashboard Studio* interface

Apama Tools

The Apama platform provides a rich, intuitive development environment that enables business analysts and IT developers to quickly and easily build, test, and evolve Apama applications. The two most important tools in the Apama platform are:

- **Apama Studio***: The Apama platform comes with Apama Studio*, a comprehensive integrated development environment (IDE) for development, debugging, testing, profiling, backtesting, and deployment of applications and services.
- **Apama Dashboards and Apama Dashboard Studio***: Further empowering users, the Apama platform incorporates dashboards to set parameters and to initiate, monitor, and terminate applications. The Apama Dashboard Studio* is a rich, intuitive, graphical interface (Figure 3) that offers layout flexibility and requires no programming to create dashboards tailored to user requirements. Dashboards can be reused, applying the same graphical display to different event processing applications.

Intel Xeon Processor E7 v3 Family Delivers Outstanding Performance, Scalability, and Reliability for Apama

To unleash the event-processing power that Apama Streaming Analytics promises, the Intel Xeon processor E7 v3 family delivers the outstanding performance, scalability, and reliability required for business-critical real-time analysis of fast-streaming big data.

Performance

The power of the Apama platform relies heavily on parallel processing. The key performance features of the Intel Xeon processor E7 v3 family that enable Apama to process relentless streams of data and turn them into insights are:

- Up to 18 cores (36 threads)
- Up to 45 MB last-level cache
- Up to 59-percent performance increase compared against previous-generation processors (see testing results in this whitepaper)
- Speed up operations with Intel® Advanced Vector Extensions 2.0 (Intel® AVX2)
- Per core P-State: Allows cores to run an individual frequency and voltage that reduces power consumption

- Cache quality of service (QoS) monitoring: Enables optimized scheduling, load balancing, and resource monitoring

Scalability

The Intel Xeon processor E7 v3 family offers outstanding scalability to enable Apama to handle any data streaming workload.

- Up to 12 TB memory support for in-memory analytics
- Flexible memory-support options (DDR4 and DDR3)
- Native scaling: 2-, 4-, and 8-socket configurations
- Extended scaling up to 256 sockets with third-party node controllers

Reliability

Real-time decision making requires systems with exceptional reliability, availability, and serviceability (RAS). The Intel Xeon processor E7 v3 family has many RAS features, including the following:

- Enhanced Machine Check Architecture (eMCA) Recovery Gen 2: channels all machine-check events through firmware first, enabling system recovery from a broader range of error conditions.
- Address-based memory mirroring: Allows mirroring of only the most critical memory regions, leaving more memory available to the system and thus reducing overall investment in memory while providing protection for critical regions.
- Multiple-rank sparing: Further reduces the frequency of server maintenance or downtime by providing additional runtime-failover options for memory ranks encountering errors beyond a set IT threshold.
- DDR4 recovery for command and parity errors: Reduces the frequency of fatal memory-error conditions that would cause system crashes.

Exceptional Throughput with Intel® Ethernet Controllers

The Intel Xeon processor E7 v3 family helps you process big data at lightning speeds. However, processors don't tell the full story. Slower network controllers can create bottlenecks that limit your overall transaction throughput, which could waste the full potential of your Intel Xeon processor. You can reduce or eliminate those bottlenecks by using 40GbE Intel Ethernet Controllers. These high-speed controllers can offer a great return on investment (ROI) for your Apama platform. For an increase of less than $150 per system, you can upgrade from 10GbE to 40GbE and benefit from accelerated bandwidth for real-time analytics, with lower latency and reduced power consumption.[7] As we'll see in the next section, upgrading the network controller from 10GbE to 40GbE can nearly double your overall throughputs.

Streaming Analytics in Action

There are many potential uses for analytics from large data streams in motion. Let's take a look at one of them: location-aware advertising for the retail industry. Most retailers have a wealth of customer information stored in their databases including purchase history, shopping patterns, and preferences. With access to a customer's geographic location, a retailer with a powerful event-processing system like Apama can quickly recognize when a customer is near or even inside one of the retailer's stores, retrieve pertinent customer information, compare it to the retailer's current offerings, and send the customer a customized alert, coupon, offer, or promotion all within seconds.

A large retailer can have hundreds of thousands of stores and millions of customers spread out across continents. With that many customers and stores, pinpointing a customer that is near or in a store at any moment might seem impossible. But the Apama platform running on a server powered by the Intel Xeon processor E7 v3 family has the power to potentially monitor the proximity of each of the 319 million people in the U.S. to a million stores every 60 seconds.[8]

Throughput Tests of Apama Running on the Intel Xeon Processor E7 Family: Internal Tests

Internal throughput tests of the Apama system performed by Software AG and Intel show an approximately 59-percent higher throughput with Apama running on a server powered by the Intel Xeon processor E7 v3 family compared to when running on a previous-generation processor.1 Software AG and Intel tested the internal throughput of the Apama Correlator in separate tests running on one server powered by the Intel Xeon processor E7 v3 and another server powered by the Intel Xeon processor E7 v2 family. These tests measured the throughput of the Apama platform for a set of producer/consumer contexts. Each producer context "published" events to an event channel that only the consumer is subscribed to. The number of producer/consumer pairs (contexts) was incrementally increased to investigate how the total throughput varied as the degree of parallelism (number of contexts) was increased. The throughput for each scenario (from one context pair up to 200 context pairs) was measured in number of events per second. The results of the tests are shown in Figure 4.

As expected, throughput increased as more contexts were added, exploiting the available parallel processing capacity of the system.

Above 20 to 25 contexts, the contexts began sharing cores through hyperthreading. At this point, throughput continued to increase

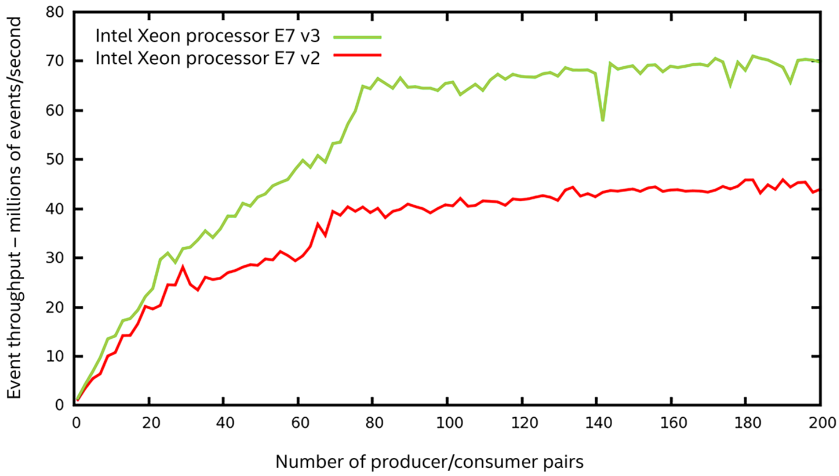

Figure 4. Internal event throughput of the Apama Correlator* as a function of producer/consumer context pairs

but at a lower rate. Between about 65 and 75 contexts, the throughput increased more sharply, which might be attributable to Apama heuristics that favor throughput over latency when the system is loaded with a high number of simultaneously running context. Finally, between about 70 and 80 contexts the system becomes saturated, with only minor increases in throughput as contexts were added.

Throughput Tests of Apama Running on the Intel Xeon Processor E7 Family: External Tests

In a second series of tests performed by Software AG and Intel, events were sent across a network to Apama Correlator running on servers powered by Intel processors. Throughput was measured on both 10GbE and 40GbE networks, with Apama running on one server powered by the Intel Xeon processor E7 v3 family and another server powered by the Intel Xeon processor E7 v2 family.2

The tests showed significant gains from upgrading the network controller from 10GbE to 40GbE. The faster network eased bandwidth constraints that were limiting throughput. Upgrading the network resulted in 1.93 times faster throughput; however, the biggest gains resulted from the combined effect of using a 40GbE network controller in combination with a server powered by the Intel Xeon processor E7 v3 family. The tests showed 2.64 times faster throughput with Apama running on a server powered by the Intel Xeon processor E7 v3 family and using a 40GbE network controller compared to Apama running on a server powered by a previous-generation processor and using a 10GbE controller (Figure 5).

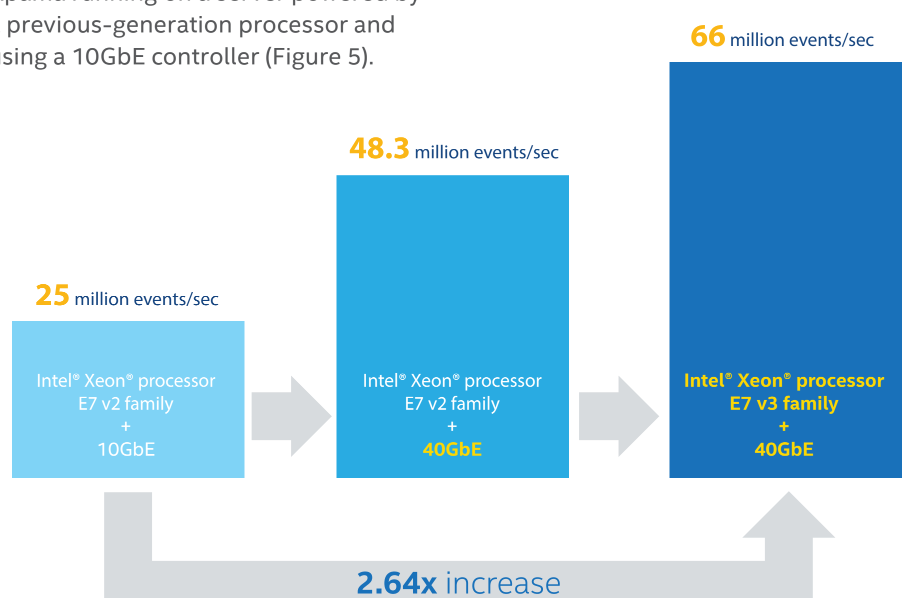

Figure 5: Event throughput for Software AG Apama Streaming Analytics* running on the current-generation Intel Xeon processor family versus the previous generation Intel Xeon processor family over both 10GbE and 40GbE networks

Conclusion

Streaming data conceals insights that can lead to valuable business opportunities and help identify potential liabilities. To reveal those insights, enterprises need a data streaming analytics platform that supports real-time performance at scale. The Apama platform, running on servers powered by the Intel Xeon processor E7 v3 family, meets these enterprise needs and expands the range of what is possible with substantial event processing performance powered by the Intel Xeon processor E7 v3 family.

To learn more how Intel can accelerate your streaming data and event-processing solutions, visit: **www.intel.com/bigdata**.

For more information about Apama Streaming Analytics, visit: **www.softwareag.com/apama**.

[1] O1 One server was powered by a four-socket Intel® Xeon® processor E7-8890 v3 and another server with a four-socket Intel Xeon processor E7-4890 v2. Each server was configured with 512 GB DDR4 DRAM, Red Hat Enterprise Linux 6.5*, and Apama 5.2*. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

[2] One server was powered by a four-socket Intel® Xeon® processor E7-8890 v3; another server was powered by a two-socket Intel Xeon processor E5-2680 v2. The server powered by the Intel Xeon processor E7 v3 family was configured with 512 GB DDR4 DRAM; the server powered by the Intel Xeon processor E5 v2 family was configured with 256 GB DDR3 DRAM. Both servers were configured with Red Hat Enterprise Linux* (RHEL*) 6.5 and Software AG Apama Streaming Analytics 5.2*.

[3] McKinsey & Company. "All Things Online." September 2013. http://www.mckinsey.com/insights/mgi/in_the_news/all_things_online.

[4] Forrester Research. "The Forrester Wave: Big Data Streaming Analytics Platforms, Q3 2014." https://info.softwareag.com/apama-leader-forrester-wave-big-datastreaming-analytics-solution.html.

[5] Software AG, 2013. "The Apama Platform." February 2014. https://www.softwareag.com/corporate/images/sec_SAG_The_Apama_Platform_WP_Feb14_WEB_tcm16-113796.pdf

[6] On Linux* 64-bit platforms as of April 2013.

[7] Prices are based on Intel's recommended customer price.

[8] Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at intel.com.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

 Printed in USA 0515/SA/PRW/PDF Please Recycle 332395-001US