White Paper

Understanding Data Streams in IoT

Contents

Author
Frédéric Combaneyre
Sr. Business Solutions Manager, SAS

The Internet of Things

The large array of connected devices, often referred to as the "Internet of Things" (IoT), is delivering an array of new data from the sensors they contain. This data offers the promise of new services, improved efficiency and, possibly, more competitive business models.

While there may be many affects from this set of communicating technologies, one thing is certain: Connected devices will result in a dramatic shift in how we all interact with computers. First and foremost, computers, computing devices and applications will surround us in an environment where the physical and virtual worlds are constantly connected. Instead of today's billion-node Internet network, the Internet of the near future will be used by trillions of devices, people, organizations and places. And a trillion-node network poses design challenges along with opportunities. Finding a way to process this information – and spot useful events – will be a distinctive factor in an IoT world.

The Early World of Sensors

The first sensors appeared decades ago, and while they have a long history, these devices entered the popular nomenclature more recently thanks to the Internet of Things. A sensor detects events – or changes in quantities – and provides a corresponding output, generally as an electrical or optical signal.

Today, sensors are used in everyday objects such as touch-sensitive elevator buttons and lamps that dim or brighten by touching the base. Sensors are also heavily used in manufacturing, medicine, robotics, cars, airplanes and aerospace.

The largest sensor challenges occur once signals have been detected. At that point, you have to decide:

- Where do I collect the data being generated?
- How can I use it?

To capture and collect the signals coming from sensors, the operational historian emerged. The operational historian refers to a database software application that logs and stores historical time-based data that flows from sensors. These data stores are optimized for time-dependent analysis and are designed to answer questions such as, "What was today's standard deviation from hourly unit production?"

Historian technology captures data from sensors and other real-time systems, and it often uses manufacturing standards and contains interfaces for several hundreds of sensor types. These dedicated data historians are also designed to survive harsh conditions, such as a production floor, and they feature the ability to continue capturing and storing data even if the main data store is unavailable.

Historian software technologies have developed complementary tools to provide reporting and monitoring features to detect trends or correlations that indicate a problem. They can then alert an operator to take immediate action before equipment damage occurs.

Until recently, that was the state of the art for generating value out of sensor data.

The Internet of Things – and Big Data Explosion

Since 2012, two major changes have shaken the sensor world – and caused the IoT market to mature more rapidly than before:

- **Sensors shrank**. Technological improvements created microscopic scale sensors, leading to the use of technologies like like Mircroelectromechanical systems (MEMS). This meant that sensors were now small enough to be embedded into unique places like clothing or other materials.
- **Communications improved**. Wireless connectivity and communication technologies have improved to the point that nearly every type of electronic equipment can provide wireless data connectivity. This has allowed sensors, embedded in connected devices, to send and receive data over the network.

Whether it's a car or a pacemaker, the data from sensors is flowing in a constant stream from the device to the network – and sometimes back to the device. This is leading to massive amounts of data, and as a result, IoT is seen as a major contributor to the era of big data.

While organizations today are investing heavily in capturing and storing as much data as possible, the critical challenge is using this data when it is still in motion – and extracting valuable information from it. Organizations are (or will soon be) scrambling to apply analytics to these streams of data before the data is stored for post-event analysis. Why? Because you need to detect patterns and anomalies while they are occurring, in motion, in order to have a considerable impact on the event outcome.

For example, monitoring live trends in traffic flow can help make traffic controls more effective, particularly with control-to-car communication instructing ways to avoid bottlenecks or even potential accidents. And sensors within household appliances can alert people to potential problems – like a refrigerator that isn't cooling properly – or even make helpful suggestions (like telling you that you'll be out of milk in less than a day at current consumption rates).

Exploiting the Internet of Things

The IoT promises to create a highly efficient world that demands constant analysis of the state of events based on the sensor and machine communications happening all around us. It's important to not only consider the rapidly increasingly amount and types of data that are created automatically, but also the exploding number of ways that data needs to be filtered, mashed up, compared, contrasted, interpolated and extrapolated. Speed (the ability to process data as fast as possible) and volume (the ability to process massive amounts of data) are now the two required qualities of any system that needs to process IoT data.

When it comes to exploiting the Internet of Things, it's useful to compare to the traditional analysis where analytics is applied on stored data – just as you would use analytical processing of events that occurred in the past, and as with the operational historian of sensors of the past. Within the IoT, however, the emphasis is to identify and examine patterns of interest as the events occur, and before the data is stored, enabling real-time actions on those events of interest. The new goal is to analyze the data as close to the occurrence of the event as possible, before any value is lost.

Industries have begun to capture sensor data and apply analytical insights in real time. Smart cities are arguably one of the biggest sources for use of monitored sensor data. With analytical models applied in real time, such infrastructure realizes:

- Improved efficiency of urban systems, like detecting traffic congestions or accidents – and instantaneously optimizing the transportation flow while sending new instructions to city light grids and alerts to commuters to allow them to take appropriate action.
- Optimized grid power networks that choose the best power source based on existing conditions and projected needs.
- Monitored water systems that prevent infrastructure failures, alert staff about leakages and help understand the impact of water usage on the surrounding environment.
- Managed infrastructure such as street lighting management, parking management, real estate and space optimization as well as public safety and security.

In the utilities industry, there is widespread use of sensors to capture every potential piece of information to help achieve 24/7 uptime. Therefore, it's critical to detect when expensive and strategic equipment is about to fail – taking corrective action before it fails.

Retailers need to optimize the shopping experience in order to increase revenue and market share. One way that is being done is with more effective and personalized consumer communications. For example, sensors are being used to detect in-store behavior. That streaming data is being analyzed, along with other store information (like inventory, social media chatter and online-shop user profiles), to send customized and personal offers while the purchase decision is underway.

For these examples, and in many others, analyzing data after it's been stored isn't sufficient – it takes too long to react to the sensed information at hand. SAS provides a suite of high-performance analytics solutions that dramatically speed the exploitation and use of the massive amounts of data held in big data stores. For real-time decision making on streaming data, it is important to have a technology that can process and provide deep analysis on data at a very high throughput speed. And for that, specialized technology that addresses complex event processing, streaming analytics – like SAS® Event Stream Processing – is required.

What is SAS® Event Stream Processing?

SAS Event Stream Processing helps you quickly decipher and analyze volumes of continuously flowing events held in streams of data. The technology examines and assesses events while they are in motion, also known as "event streams." Instead of running queries against stored data, the technology stores the data management and analytical routines and streams massive amounts of data through these queries – filtering, normalizing, aggregating the data and detecting patterns in real time. This process occurs before the data is stored to reduce the latency of the insights derived, influencing actions in real time.

SAS Event Stream Processing is an embeddable engine that can be integrated into other SAS solutions. As the data is continuously processed and analyzed, it is streamed to other SAS applications (such as SAS Asset Performance Analytics, SAS Decision Manager, SAS High-Performance Risk) for in-depth analysis. SAS Event Stream Processing initiates and executes

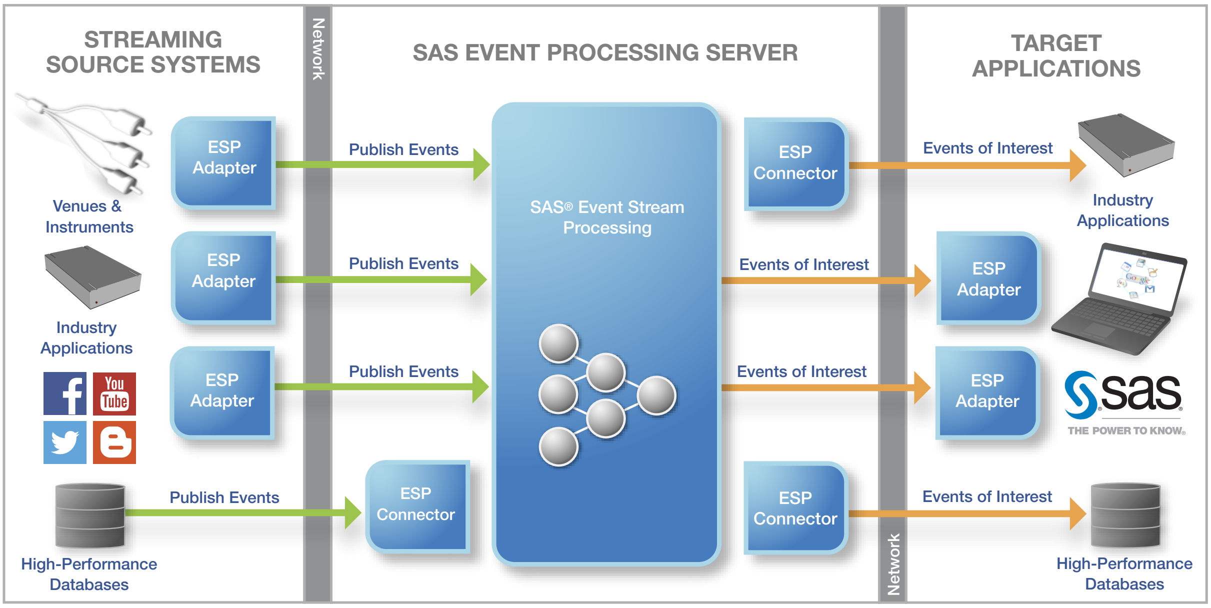

Figure 1. Conceptual view of SAS Event Stream Processing architecture.

a predefined query against incoming events in real time and in memory. Events are kept in memory, and a retention period can be defined that instructs the depth of history to be held in memory. This allows the engine to perform more complex data manipulation and pattern detection for a rolling period of time that is of interest for the specific event(s).

SAS Event Stream Processing uses the following techniques to manage, and make sense of, streaming data:

- **Assessment**. The volume of data generated by sensors may make it impractical to store it all. In fact, much of what is generated may be deemed “irrelevant” without need for further analysis or archiving. Event stream processing helps you to standardize the data as it arrives, applying simple transformations and rules to determine if further downstream processing needs to occur. If not, the data (or event) can be quickly discarded, without taking up additional processing bandwidth.
- **Aggregation**. Let’s say you wanted to detect fraudulent gift card use. Your business rule could be: “Tell me when the value of gift card redemptions at any point-of-sale (POS) machine is more than $2,000 in an hour.” Event stream processing can continuously calculate the necessary metrics – across sliding time windows – to understand real-time trends in gift card redemption. This kind of continuous aggregation would be difficult, if not impossible, with traditional tools.
- **Correlation**. Event stream processing allows you to connect to multiple streams of data that is in motion (even over a period of time of seconds or days), and identify that a series of events occurred, e.g., condition A was followed by B, then C. For example, if you connect to streams of gift card redemptions from 1,000 POS terminals, event stream processing can continuously identify conditions that compare the different POS terminals to each other. A sample rule might be “Generate an alert if gift card redemptions in one store are more than 150 percent of the average of other stores.”
- **Temporal analysis**. Event stream processing uses the concept of time as a primary computing element, which is critical for understanding scenarios where the rate and momentum of change matters. For example, sudden surges of activity can be clues to potential fraud. Event stream processing can detect such surges as they occur. A rule here could be, “If the number of gift card sales and card activations within four hours is greater than the average number of daily activations of that store in the previous week, stop approving activations.” Unlike traditional computing models – which are designed to summarize and roll up historical data - event stream processing asks and answers these questions on data as it is changing.

These capabilities set event stream processing apart from other information approaches by revealing what’s happening now, not just what happened in the past, so you can take action immediately. It can be used to predict what will happen next, when predictive models are embedded into the data streams, and to even determine where optimizations can be made for better outcomes.

Figure 1 shows a typical architecture of the SAS Event Stream Processing in action, composed of:

- Source systems publishing events to the SAS Event Stream Processing Engine through adapters or connectors.
- An event stream processing server (UNIX or Windows) that runs one or several instances of SAS Event Stream Processing. SAS Event Stream Processing processes the events according to the defined queries and models encoded and loaded for processing execution.
- Target applications that subscribe to and receive events of interest through adapters or connectors.

Of course variations of this architecture are implemented, configured to the business needs or architecture constraints.

Event Stream Processing Meets IoT

Event stream processing technologies were initially designed to tackle "traditional" streams of data – in particular financial market trades. With connected things, streams of data are creating a new world of decision and service capabilities.

Here are a few ways that event stream processing plays a vital role in IoT:

- **Detect events of interest and trigger appropriate action**. Event stream processing pinpoints complex patterns in real time, generated by a person's behavior on their device. It can also discern the state of a piece of equipment, such as a pending system failure or an upcoming maintenance event. Event stream processing technologies are also used to detect potential fraud, opportunities to send a personalized marketing offer, as well as propagating information to dedicated systems for immediate attention.

- **Aggregate information for monitoring**. Event stream processing is also used to continuously monitor sensor data from equipment and device sources, monitoring for trends, correlations, or defined thresholds that indicate a problem – alerting an operator to take immediate action before the damage occurs.

- **Sensor data cleansing and validation**. Sensor data is notoriously "dirty." There are often missing time stamps for a single sensor, mostly related to network issues. When multiple sensors are being monitored as a collective, different formats and transmission timing often vary between sensors. As a result, sensor data can be incomplete or contain inconsistent values. Delayed data might even indicate a potential sensor failure or simply be the result of a drop in a mobile network. Utility companies, for example, have standard rules for data quality pattern detection. There are a number of techniques that can be directly embedded into data streams that address sensor data quality. All of these methods involve caching some sort of history to examine the erroneous nature of data issues.

- **Real-time predictive and optimized operations**. Processing event stream data, although a core consideration, isn't sufficient to empower real-time decision making. Streaming data must include analytical power to understand patterns that provide distinctive value. It's well recognized that the value in IoT data will come from analysis of what the current conditions are and any proactive actions that can be taken. Advanced analytics and mathematical algorithms developed using rich histories of stored data can be encoded into data streams – for continuous scoring of streaming data. In the real world, this could mean that information on a transit train's arrival could be fed through a series of calculations to determine how its arrival can affect other vehicles. Real-time calculations can help minimize the impact on travelers by delaying a train at the upcoming station so that people don't miss their connection. Or it can alert a bus driver to hold a few minutes longer to avoid stranding passengers at that station.

> With connected things, streams of data are creating a new world of decision and service capabilities.

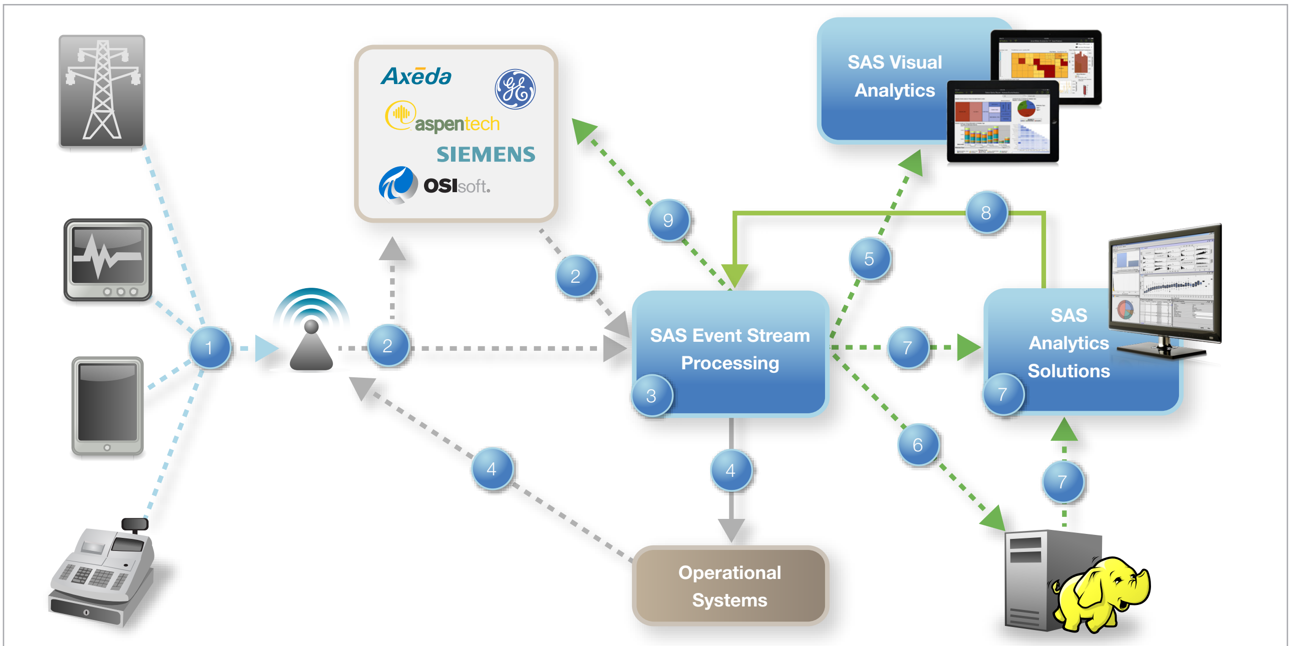

Figure 2. Conceptual view of an Internet of Things architecture using SAS Event Stream Processing.

SAS® Event Stream Processing Integrated Into the Internet of Things

SAS Event Stream Processing has worked in very complex environments that are on the early stages of what many would call the Internet of Things. Here is one scenario of how this technology works within an emerging IoT environment.

In the architecture depicted in Figure 2, there are several communication and data flowing linkages illustrated with a number system. These are:

1. Devices and connected equipment that continuously generate data from sensors, transmissions, events and human actions. The data is sent over a wireless (and physical) network, usually continuously, in real time.

2. All the data is sent to operational historian systems and cloud based aggregators that are responsible for collecting and storing this information as time-stamped data. SAS Event Stream Processing connects to these historians, receiving a continuous flow of data. For some equipment such as phone devices that are based on an operating system like Linux (and are basically small computers), the devices/equipment can directly send events to SAS Event Stream Processing (without using a bus communication system).

3. SAS Event Stream Processing processes all received data (i.e. events). This typically includes integrating different sources of streaming data as well as individual data elements, normalizing the data, cleansing it and addressing things like missing data, different formats with different rates or different transmission protocols. It also includes aggregating the data and detecting patterns of interest, using combinations of business rules, advanced analytic models and text data extractions. This transforms the detailed data streams into meaningful events and information regarding the situational status of emitting devices.

4. When SAS Event Stream Processing has detected an event, and action is required, it sends notification to a responsible person or system workflow and can trigger an operational system to perform a prescribed action. This can include stopping a machine, changing a temperature setting, activating vehicle brakes, changing streetlight intensity, sending a marketing offer or any other desired action.

5. Events can be continuously monitored for trends, correlations, calculated thresholds or other statistics. If found, the system triggers manual actions performed by authorized personnel who are focused on situational awareness assessment in real or near-real time.

6. Working in parallel with step 5, all events (or only those selected as relevant to business needs) are stored in a dedicated storage system. Due to the amount of data generated by sensors and devices, this will typically need to be a high-volume storage system, like Hadoop.

7. SAS Analytics solutions are used to understand changes in patterns or emerging events using a mix of both historical data from existing systems combined with the real-time, streaming events coming directly from the SAS Event Stream Processing. Dedicated solutions such as SAS Asset Performance Analytics, SAS Fraud Framework and SAS Intelligent Advertising can provide sophisticated, in-depth analytics to identify complex patterns of event behavior.

8. Once these complex patterns are identified with SAS Analytics, they are brought into the live, streaming environment, improving the SAS Event Stream Processing logical and analytical rules to continuously improve and keep pace with changing conditions. Steps 7 and 8 both provide a feedback mechanism to ensure ongoing value is derived from IoT data.

9. SAS Event Stream Processing is also used prior to data being stored in operational historians and cloud providers. In this use case, it's aggregating IoT data, cleaning it and normalizing dirty sensor data (e.g., across devices, sources, times) in real time. And it can also calculate desired analytical scores, sending this cleansed, enriched data back to the historian application or other data storage systems.

Conclusion

Connected, communicating devices – what we call the "Internet of Things" – are the next big data source. But it's no longer just a matter of big data. Now, it's faster data, arriving at a speed that isn't perceptible to the human eye and which can seem dizzying, especially if you're unclear how to get value from it.

Streaming data, and machine-generated data in particular, requires technology options like event stream processing to ensure optimum processing and insight. Data coming from the sensors of connected devices are the key source of streaming data now and in the future. SAS Event Stream Processing helps to acquire, understand and exploit this data supporting realtime surveillance, analytically sound actions and probabilistic analysis and alerts.

To learn more about getting value from IoT data, please visit us at sas.com/en_us/insights/big-data/internet-of-things.html.

More details about SAS Event Stream Processing are available at sas.com/en_us/software/data-management/event-stream-processing.html.

To contact your local SAS office, please visit: sas.com/offices

THE POWER TO KNOW.