WHITE PAPER

Understanding Data Streams in IoT

Contents

Devices connected through the Internet of Things (IoT) appear in many settings – cars, factories, households, retail stores and wearables, just to mention a few. Behind the scenes, these "things" stream massive amounts of data from the sensors they contain. And they do it nonstop, with a constantly changing data stream that has sparked the creation of new services, monetization opportunities, better efficiency and more competitive business models.

The data from sensors in IoT devices is used by trillions of other devices, people, organizations and places. While this trillion-node network poses many challenges, opportunities also abound. Organizations that are prepared to collect, process and analyze this barrage of IoT information can distinguish themselves from competitors in ways they might have never imagined in the past.

The Early World of Sensors

The first sensors appeared decades ago. These early sensors were designed to detect events or changes in quantities, then provide a corresponding output – generally as an electrical or optical signal. Soon appearing in everyday objects, like touch-sensitive elevator buttons and lamps that dim or brighten when you touch the base, such sensors weren't necessarily connected to each other, or to the internet. Sensors like these have been used for many purposes over the years – in manufacturing, health care, energy, robotics, cars, airplanes and aerospace.

To capture and collect the signals coming from sensors, the operational historian emerged. This database software application logs and stores historical, time-based data that flows from sensors. The data stores are optimized for time-dependent analysis, which happens after the data is stored – and they're designed to answer questions like: "What was today's standard deviation from hourly unit production?"

Historian technology often uses manufacturing standards and captures event data from hundreds of sensor types and other real-time systems. These dedicated data historians can survive harsh conditions, such as a production floor, so they continue capturing and storing data even if the main data store is unavailable. Historian software often includes complementary tools for reporting and monitoring on historic data, and can detect trends or correlations. When an issue is flagged, the system can alert an operator about the potential problem.

This used to be an advanced way to generate value out of sensor data. But with the rise of the IoT, the uses for sensors – and the data streaming from them – have become much more diverse.

From cars to pacemakers, the data from sensors flows in a constant stream from the device to the network – and sometimes back to the device. These massive amounts of data have made the IoT a major contributor to big data.

The Big Data Explosion and Sensors in the Internet of Things

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

- **Sensors shrank**. Technological improvements created microscopic scale sensors, leading to the use of technologies like microelectromechanical systems (MEMS). This made sensors small enough to be embedded into unique places like clothing.
- **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 allows sensors, embedded in connected devices, to quickly send and receive data over a network.

Today, organizations are investing heavily to capture and store as much data as possible. But their bigger challenge is to extract valuable information from the data while it's still in motion, as close as possible to the occurrence of the event. If you wait to analyze data after it's stored, it takes too long to react. That could mean missing a new business opportunity or losing out to a competitor.

In many ways, the IoT promises to create a highly efficient world. But achieving it demands constant analysis of the state of events based on sensor and machine communications happening all around us.

To take full advantage of data streams in the IoT, organizations must understand the exploding number of ways "big" IoT data needs to be filtered, mashed up, compared, contrasted, interpolated and extrapolated. Consider:

- **Volume**. Can you quickly access, integrate, store, process and analyze today's massive amounts of data?
- **Variety**. New types of IoT data are still emerging. Can you manage all the different types of data and the varied formats – structured, unstructured, semistructured – on the fly?
- **Velocity**. Think about how quickly text, image and video data is generated by cell-phone cameras, social media and devices like smart watches. That's only a small part of the data tsunami. Can you act quickly enough to capture and analyze all that data?
- **Veracity**. In its raw form, IoT data is "dirty" – it hasn't been filtered, validated, profiled or cleansed. Making IoT data trustworthy so it can be used as the basis for data-driven decision making calls for data management standards like data quality and data governance. Newer technologies, such as blockchain, can also be used to ensure the original data sources can be trusted.

The biggest challenge with sensors occurs after signals have been detected. At that point, you must decide:

- Where do I collect the data being generated?
- What do I keep and what can be discarded?
- How can I use it?

Event Stream Processing Meets IoT

Event stream processing plays a vital role in handling IoT data from the connected world. It can:

- **Detect events of interest and trigger appropriate action**. Event stream processing pinpoints complex patterns in real time. This could, for example, be generated by a person's behavior on their mobile device, or by an unusual activity during a banking transaction. Event stream processing can quickly detect potential fraudulent activities or identify prime opportunities for sending a real-time, personalized marketing offer.
- **Monitor aggregated information**. Event stream processing continuously monitors sensor data from equipment and devices, looking for trends, correlations or defined thresholds that indicate a problem. In turn, an operator can be alerted to act before damage occurs.
- **Cleanse and validate sensor data**. Sensor data is notoriously dirty. There are often missing time stamps for a single sensor due to network issues. When multiple sensors are monitored as a collective, formats and transmission timing can vary between sensors. As a result, sensor data may be incomplete or contain inconsistent values. Delayed data might indicate a potential sensor failure, or it could simply be the result of a drop in a mobile network. A variety of techniques – embedded directly into data streams – can detect patterns and examine the erroneous nature of data issues.
- **Enable real-time predictive and optimized operations**. Streaming data combined with analytics reveals patterns that empower real-time decisions. Advanced analytics and mathematical algorithms are developed using rich histories of stored data that can be encoded into data streams, enabling continuous scoring of streaming data. So, for example, information on a transit train's arrival could be fed through a series of calculations to determine how its arrival will affect other vehicles. Real-time calculations could minimize the impact on travelers by delaying a train at the upcoming station, so people won't miss their connections.

Some Industry Examples

Opportunities to capture sensor data from the IoT and apply analytical insights to it in real time are diverse, crossing many different industries. Following are a few examples.

Smart Cities

Smart city infrastructures monitor a vast array of sensor data. By applying analytical models in real time to IoT sensor data, a city can:

- Improve efficiency of urban systems, such as highways and traffic lights. For example, the IoT infrastructure could detect traffic congestion then instantaneously optimize the transportation flow by sending special instructions to city light grids and alerting commuters to take alternate routes.
- Optimize grid power networks by choosing the best power source based on existing conditions and projected needs.
- Monitor water systems to prevent failures, alert staff about leaks and better understand the impact of water usage on the surrounding environment.
- Manage infrastructure such as street light management, parking management, real estate and space optimization, and public safety and security.

Utilities

The electric utility industry makes widespread use of sensors to capture every potential piece of information – from generation to transmission, distribution and retail. This data helps utility companies work toward 24/7 uptime for customers. IoT analytics can detect underperforming assets and predict problems before expensive equipment fails. IoT technology also helps utilities integrate distributed energy sources, such as solar and wind. And it helps address disruption in the industry and long-term revenue growth by capturing and analyzing data from third-party providers and energy customers.

Retailers

No one wants to have a totally different experience from the same retailer when they're shopping online and in the store. Retailers can use IoT analytics to build a better omnichannel experience for customers – leading to an increase in revenue and market share. For example, retailers can use WiFi, beacon and RFID technologies to detect in-store behavior. When they combine and analyze this type of streaming data along with other information (like inventory, social media chatter and online-shop user profiles), they can send personalized offers while a purchase decision is underway.

SAS® Event Stream Processing: Managing and Making Sense of Data Streams

To make real-time decisions on streaming data, it's important to process and provide deep analysis on data at a very high throughput speed. SAS Event Stream Processing can quickly decipher and analyze high volumes of continuously flowing events held in data streams. Instead of running queries against stored data, it stores 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 approach reduces latency of the insights derived and influences actions in real time.

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

- **Ingestion**. Adaptors ingest data streams from a variety of sources at or behind the edge. This is the first step in a streaming analytics architecture that collects data from diverse sources to analyze it for patterns and other insights.
- **Assessment**. It's impractical to store all the data that's generated by sensors, especially since much of the data is irrelevant. Event stream processing can standardize the data as it arrives, applying simple transformations and rules to determine if additional downstream processing is needed. If not, the data (or event) can be quickly discarded, without taking up additional processing bandwidth.
- **Aggregation**. Imagine that you want to detect fraudulent gift card use. You could set a business rule that says, "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.
- **Correlation**. Event stream processing lets you connect to multiple streams of data in motion – over a period of seconds, minutes or days – to see that a series of events occurred. For example, you could see that condition A was followed by B, then C. If you connect to streams of gift card redemptions from 1,000 POS terminals, for example, 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 matter. For example, sudden surges of activity can be clues to potential fraud. Event stream processing can detect such surges as they occur. A rule 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.

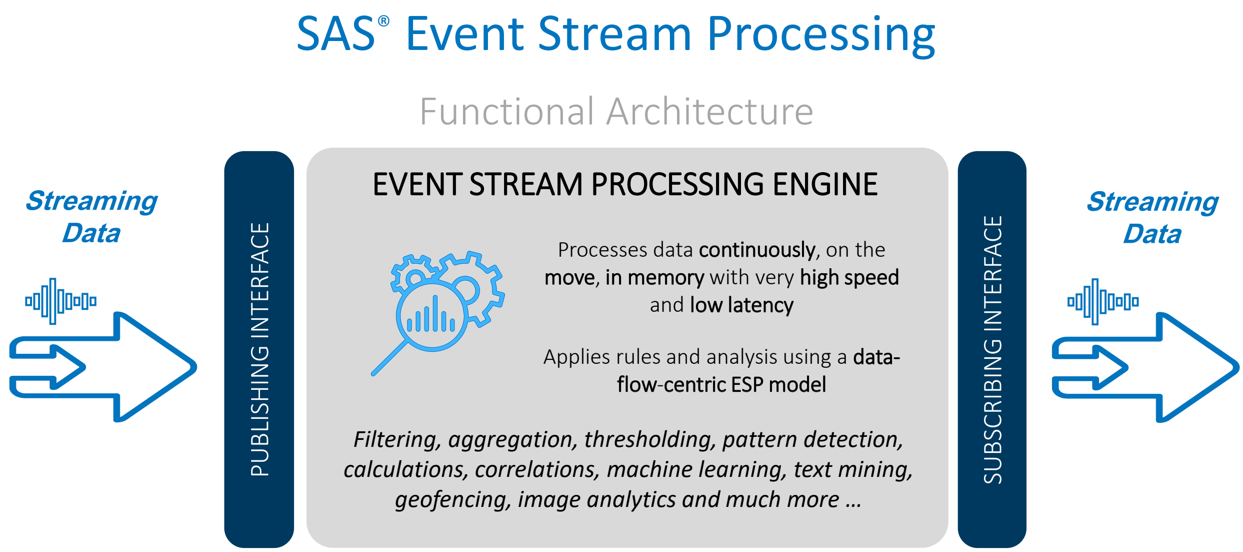

Figure 1: A conceptual view of SAS Event Stream Processing architecture, which works as follows:

- Source systems publish events to the SAS Event Stream Processing engine through streaming data adapters that ingest events from high-speed sources.
- An event stream processing server runs one or several instances of SAS Event Stream Processing, which processes events according to the defined queries and models encoded and loaded for execution.
- Target applications subscribe to and receive events of interest through adapters – so they can listen for events of interest to be acted upon in real time.

SAS® Event Stream Processing in Action

SAS Event Stream Processing is an embeddable engine that integrates with other SAS software. As data is continuously processed and analyzed, it's streamed to other SAS applications for in-depth analysis – such as SAS Quality Analytic Suite, SAS Customer Intelligence, SAS Decision Manager or SAS Risk Management. The following diagram depicts how an IoT architecture using SAS Event Stream Processing works.

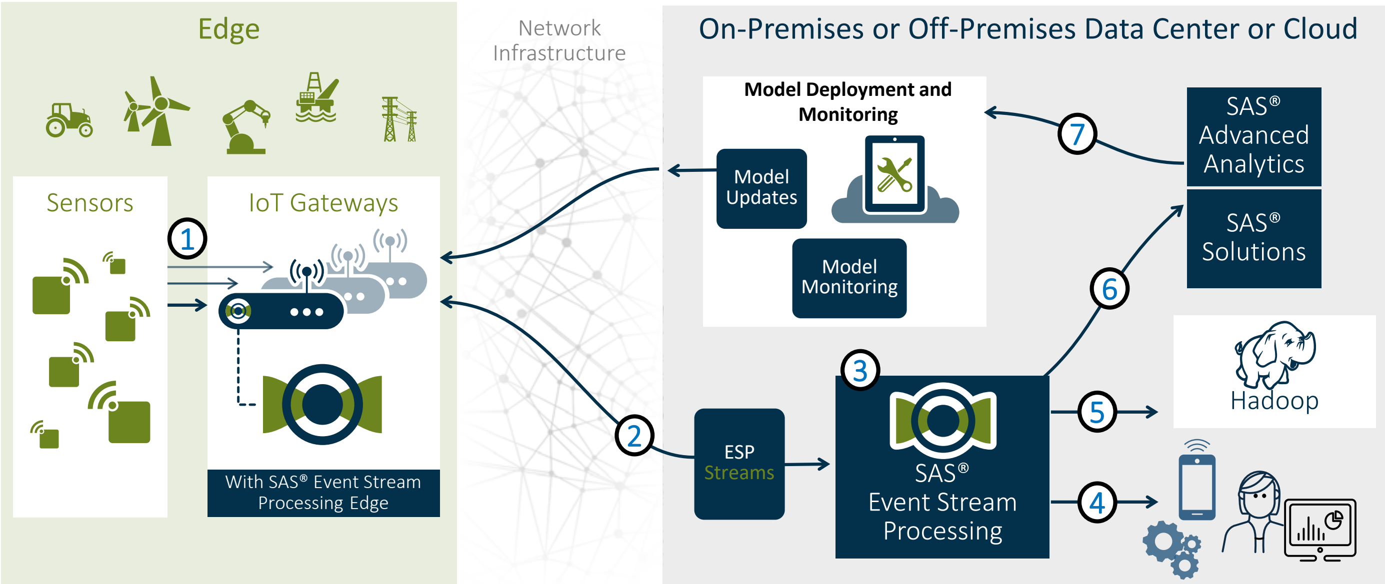

Figure 2: An Internet of Things architecture using SAS Event Stream Processing.

The architecture depicted in Figure 2 illustrates the following steps:

1. Devices and connected equipment continuously generate data from sensors, transmissions, events and human actions. This data is sent in real time over a network, usually continuously, and is processed on IoT gateways at the edge. The data can be analyzed at the edge – close to the data source – and is intelligently filtered so that only the necessary data is transmitted back to the centralized compute facility where additional SAS Event Stream Processing engines are analyzing the incoming data.

2. All the data is sent to operational historian systems and/or cloud-based aggregators that collect and store this information as time-stamped data. SAS Event Stream Processing connects to these historians or aggregators, receiving a continuous flow of data.

3. SAS Event Stream Processing processes all received data, or events. This includes integrating different sources of streaming data and individual data elements, normalizing the data, cleansing it and addressing issues like missing data, different formats with different rates, or different transmission protocols. It also aggregates the data and detects patterns of interest using combinations of business rules, advanced analytics models and text data extractions. This transforms the detailed data streams into meaningful events and information about the status of emitting devices.

4. When SAS Event Stream Processing has detected an event and action is required, a responsible person or system workflow subscribes to events of interest and triggers 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, etc. Events can be continuously monitored for trends, correlations, calculated thresholds or other statistics. If found, the system can trigger manual actions to be performed by an authorized person.

5. Working in parallel with step 4, all events (or only those selected as relevant to business needs) are streamed and stored in a dedicated storage system. Due to the amount of data generated by sensors and devices, this typically needs to be a high-volume storage system like Hadoop.

6. To understand changes in patterns or emerging events, SAS uses a mix of historical data from existing systems combined with real-time, streaming events coming directly from the SAS Event Stream Processing engine. Dedicated solutions, such as SAS Quality Analytic Suite, SAS Fraud Management and SAS Customer Intelligence 360, can provide sophisticated, in-depth analytics to identify complex patterns of event behavior.

7. As complex patterns are identified, they're brought into the live, streaming environment to continuously improve the logical and analytical rules of SAS Event Stream Processing. Streaming models deliver real-time insights and enable the software to keep pace with changing conditions by combining analytical models, business rules and various transformations of streaming data. SAS Event Stream Processing also provides a feedback mechanism to help you derive ongoing value from IoT data.

Sense, Understand and Act on IoT Data

SAS Event Stream Processing is core to the SAS solution for helping you get full value from the rich data streaming from the IoT. SAS Analytics for IoT (which includes SAS Event Stream Processing) integrates streaming data with analytics and visualization to cover the full analytics life cycle. Built on a scalable, edge-to-enterprise platform that bridges IT and operational environments, SAS Analytics for IoT enables you to:

- **Sense what matters**. By performing real-time data management on IoT event data, filtering signals from noise, you can stay focused on what's relevant for your business.
- **Understand the signals in data**. By mining and analyzing IoT data throughout the connected ecosystem – and combining it with data from other sources that add context – you can detect patterns of interest as events occur and understand what they mean.
- **Act with speed and confidence**. By collaborating and deploying analytics consistently across the IoT analytics life cycle – from accessing and preparing data to doing exploratory analysis, building and comparing machine learning models, and implementing predictive models – you'll be prepared to take the action that's best for your business.

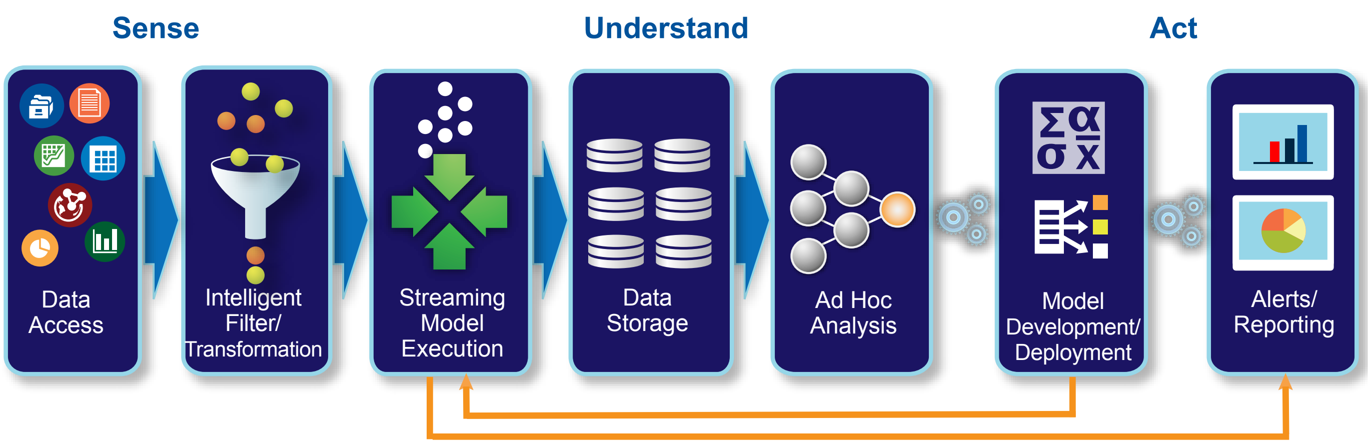

Figure 3: To achieve value from the connected world, you first need to access IoT data to sense what's important. Next, you need to understand (via analytics) what the data is communicating so you can gain insight from it. Finally, you must act quickly to get results – by developing new value chains and business processes that are responsive to the connected world.

Why SAS?

As a technology leader, you need an innovative strategy that can drive digital transformation through the IoT. SAS can reduce risk and speed time to value through our breadth and depth of analytics – including artificial intelligence – as well as our diverse interoperability, strong partner ecosystem and deep domain expertise. We have a proven track record at helping organizations create and sustain value from IoT data and initiatives, and analysts consistently rank the SAS Analytics platform as a market leader.

Learn More

Discover why SAS is an ideal partner to help you get more value from massive volumes of IoT data streams in the connected world. To learn more, visit: sas.com/iotsolutions.

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

THE POWER TO KNOW.