Smartmeter Case Study for SAS®/Intel® Solutions for Internet of Things

Suleyman Sair, Intel Corporation, Chandler, AZ

Pat Herbert, SAS Institute, Cary, NC

ABSTRACT

The amount and quality of data generated in Big Data usage scenarios present a big challenge to those looking to benefit from such data. We present a smartmeter use case that shows how one can collect data from the edge of the network and potentially aggregating and transforming the data before shipping them to the data center for in depth analysis and report generation. On the edge devices, we show how such data can be collected in a smart fashion to reduce the quantity and improve the quality of the data being collected with SAS Event Stream Processing. When employed, intelligent gateways can collect edge device data and act as a filter to summarize and normalize the data for sharing through the network and into the cloud. This information then gets sent to the data center for advanced and in depth analytics. The customizable reports generated at the data center can be viewed anywhere utilizing mobile technologies. Our use case demonstrates how software and hardware technologies from SAS and Intel work together to create solutions that give developers and customers flexibility, interoperability, security, and scale to help drive market adoption.

INTRODUCTION

Smart grid technology has revolutionized how electric utilities conduct their business. Forecasting consumption has always been a critical activity for utility companies since electricity cannot be stored and must be produced on demand. With smart grid technology, utilities have access to vast amounts of consumption and location data. In addition, coupled with weather station data and weather forecasts, there is an opportunity to improve the accuracy of the forecasts by taking weather conditions into account. Furthermore, data from transformers and substations provide more information that can be used to predict failures and prevent costly outages. And finally, advances in software and hardware allow for samples to be collected more frequently to respond to events as close as possible to the moment that they happen.

Taking advantage of all this data requires an advanced analytics platform. SAS Analytics software and Intel hardware components provide the ideal building blocks for such a platform. On the edge, SAS Event Stream Processing running on Intel Moon Island gateway devices provide the means of gathering information. Actions can be taken based on this streaming data at the edge based on event models programmed into the gateways. In addition, either the raw streaming data or a summary of that data generated at the gateway is passed on to a cloud analytics backend for richer analytics. We use SAS Visual Analytics and Intel E7 v3 processors as the foundational pieces of this cloud backend architecture.

In this paper, we describe the components and architecture of the platform we created for this study. Along with this platform, we use an enriched data set based on real smartmeter and transformer data to look at implications of predicting failures and forecasting demand. Data from two different customer data sources were combined and also the data was anonymized. Some additional data was propagated/manufactured by SAS just to generate more data and a richer data set.

ROLE OF ANALYTICS FOR UTILITIES

Many utilities already use forecasting and prediction to address their production and maintenance challenges, but they will increase in importance and relevance with the availability of more data inputs from a data-rich smart grid environment.

Forecasting is a data-intensive application that utilities use for planning, investment and decision-making. It attempts to determine how customers will use energy and then plan utilities' operations around that possible use. In addition, forecasts try to predict how customers respond to prices, and changes in weather at regular intervals. Furthermore, location data is also important as not all customers in the same territory are affected by weather in the same way. Temperature alone is not sufficient in forecasting and more factors affecting load must be taken into consideration. With improved technology, data collection from more data sources on a more frequent basis is made possible, which in turn leads to more accurate forecasts.

In short, forecasters are facing new challenges such as how to manage large volumes of energy usage data from the smart grid, build forecast models for ever decreasing time increments, e.g. every hour or every 15 minutes, and identify which new indicators impact the forecast and how to quantify their relevance.

INTEL® HARDWARE COMPONENTS

At a high level, the hardware infrastructure consists of components on the edge of the cloud and the analytics backend. At the edge, sensors embedded in smartmeters and transformers periodically send information to gateways stationed at power substations. The gateways run data quality, transformation, and filtering/aggregation procedures before sending the aggregated or the raw data to the analytics servers in a public or private cloud backend.

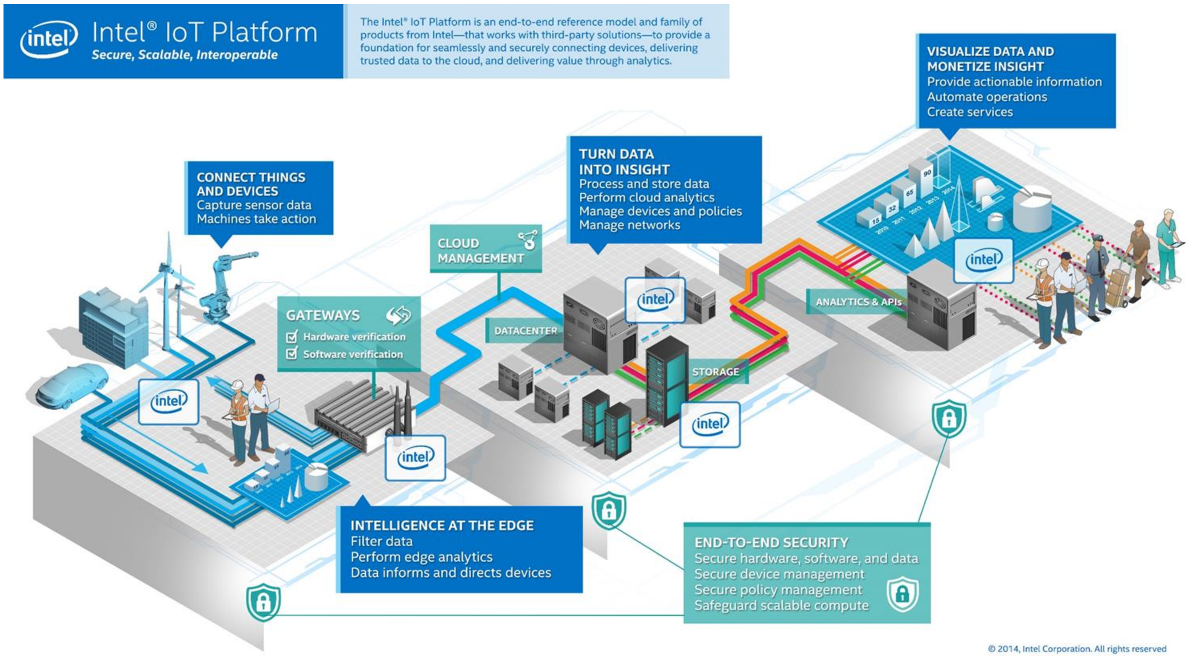


Figure 1. Caption for Sample Figure

INTEL MOON ISLAND GATEWAYS

A basic gateway features two or more interfaces:

1. Interface(s) to devices and sensors to collect data
2. A connection to a LAN, Cellular, or Wi-Fi network that will transmit data to the cloud

Gateways need to provide security features like proper authentication of the devices and services, as well as data encryption. Gateway devices typically support a broad range of connectivity protocols to satisfy vertical specific requirements such as Ethernet or ZigBee etc. Gateways can perform edge analytics to act on streaming events near the source to improve response times. They can improve the quality of the data before it is transmitted to the backend (e.g. filter bad data out or fill in missing data). Summarizing the data is another typical function performed in the gateways.

Intel© has gateway solutions based on Quark™, Atom ™, and Core™ processors available to developers. These platforms provide various performance/power/size tradeoffs to fit the needs of different applications.

INTEL® SERVER COMPONENTS

Intel® Xeon® processors are commonly used in the analytics backend to perform the advanced analytics functionality in a private or public cloud.

Today's advanced analytics solutions are fundamentally changing the speed with which businesses can extract insights from large data sets. With results delivered in seconds or minutes rather than hours or days, business intelligence can be integrated into critical, real-time processes to drive better results, improve decision making, deliver new services and experiences, and implement new revenue-generating business models.

Four-socket, eight-socket, and larger servers based on the Intel® Xeon® processor E7-8800/4400 v3 product families provide exceptional performance and scalability for real-time analytics operating on multi-terabyte and even petabyte scale datasets. With their massive execution resources, large memory capacity, and advanced reliability features, these servers also provide best-in-class support for traditional enterprise applications, such as enterprise resource planning (ERP), data warehousing, and online transaction processing (OLTP).

For building out large node count clusters with two-socket servers, Intel® Xeon® Processor E5-2600 v3 product family delivers significant performance improvements with up to 18 cores per socket, 45 MB of last-level cache (LLC), and next generation DDR4 memory support in analytics workloads.

Complementing the server platforms, Intel® Atom® and Core® based tablets provide the ideal mobile devices to view analytics reports created in SAS Visual Analytics for on-the-go access to reports that are generated from pre-created templates using real-time data.

Additional components, such as the Intel® Ethernet Controller XL710 Series, delivers proven 10 and 40 Gigabit Ethernet connectivity for the platform, extending Intel® Virtualization technologies beyond server virtualization to network virtualization. And finally, Intel® Solid-State Drive Data Center Family for PCIe* built on the NVMe* specification, is a comprehensive product family of 2.5-inch and Add-In-Card form-factors to deliver breakthrough performance that is optimized for applications such as SAS analytics solutions.

SAS® SOFTWARE COMPONENTS

The main ingredients to a solution for smartmeter IoT analytics are SAS® Event Stream Processing and SAS® Visual Analytics.

SAS® Event Stream Processing, includes a server, which provides the run-time environment for executing the event stream models against one or more data streams. The publishing interface, the Event Stream Processing Server, and the subscriber interface comprise the three main components of the SAS® Event Stream Processing engine.

Connectors in the publishing interface publish the event streams into source windows. Published operations natively read event data from a specified source and place that event data in a source window of a running event stream processor.

The SAS® Event Stream Processing server applies an event stream model that specifies how streaming data is to be transformed for the consumption of the output location subscribers. The server supports running continuous queries such as pattern definition, transformations and analytics that are applied to the data.

A subscriber interface shares the results with other systems or information consumers. Adapters and connectors can be subscribed to event stream windows. The subscribed operations natively write output events from a running event stream processor window to the specified target.

SAS® Event Stream Processing applications can analyze millions of events per second, with latencies in the milliseconds – even on small, commodity servers with 4-8 cores. Because of the lightweight architecture and thread pool management, it is easy to linearly scale the environment to run multiple SAS Event Stream Processing servers in a grid to distribute the load and extend in-memory resources.

SAS® Visual Analytics provides a complete platform for analytics visualization, enabling customers to identify patterns and relationships in data that weren't initially evident. Interactive, self-service BI and reporting capabilities are combined with out-of-the-box advanced analytics so everyone can discover insights from any size and type of data, including text. SAS® Visual Analytics provides an interactive user experience that combines advanced data visualization, an easy-to-use interface and powerful in-memory technology. This lets a wide variety of users visually explore data, execute analytics and understand what data means. Then they can create and deliver reports wherever needed via the web, mobile devices or Microsoft Office applications

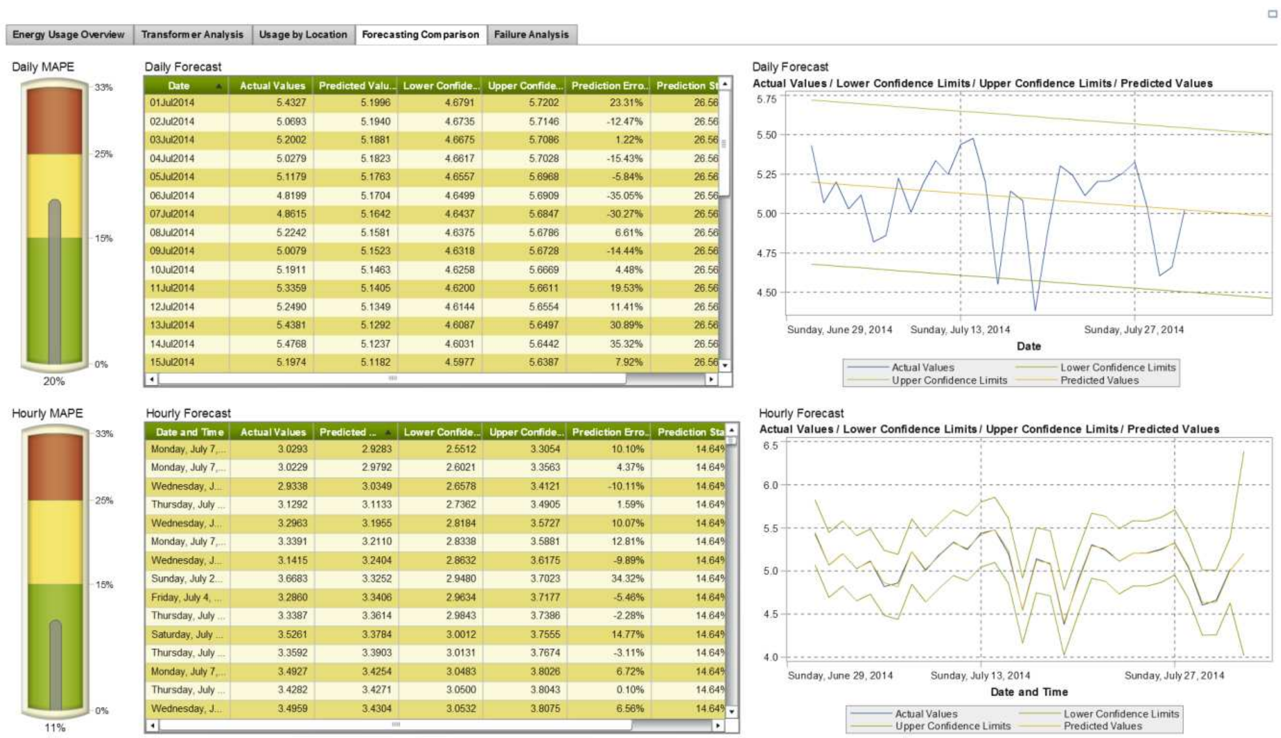
Along with SAS® Event Stream Processing and SAS ® Visual Analytics, a combination of the following SAS® 9.4 components are typically installed based on customer needs:

- SAS® BASE
- SAS® Grid Manager
- SAS® HPA Suite
- SAS® LASR Suite
- SAS® In-Memory Statistics for Hadoop, SAS® Visual Analytics, SAS® Visual Statistics
- SAS® Data Management Advanced
- SAS® Federation Server
- SAS® Event Stream Processing Engine, SAS® Real-time Decision Manager

SAS® VISUAL ANALYTICS TEST RESULTS

We used anonymized sample data from smart meters and transformers from various utilities and combined them together to increase the volume of the data. This corresponds to data from 7 cities and 38000 smartmeters. Display shows a comparison of building a forecast model with using daily consumption data when compared to using hourly consumption data. The top half of the graph shows the forecast using the daily data. The blue line is the actual usage throughout the day – with all the peaks and valleys, but the forecast is the straight yellow line – a very basic approximation of the day's usage. In the bottom half of the graph, we show the same forecast using hourly data. Again, the blue line is actual usage and the yellow line is the hourly forecast. Notice how tightly the forecast follows the actual usage. The processing capabilities in the Intel® Xeon E5 and E7 platforms enable us to extract this extra accuracy from our consumption forecasts. In this case the MAPE improves from 20% to 11%.

We also compared the accuracy of predicting transformer failures based on the consumption forecast, weather forecast, and the age and the model of the transformer. Display 2 shows the results. We have split the predictions for above ground and underground transformers since they require different forecast models as can be observed in the results. The underground transformers are less sensitive to weather variances and there are fewer data points for those types. As a result of this, the forecast model used to accurately predict above ground transformer failures does not perform well on underground transformers. However, the model is very accurate for above ground transformers, predicting failures with 96% accuracy. Correct predictions lead to substantial savings in maintenance costs as well as operating costs by preventing unscheduled outages.



Display 1. Consumption forecasts using daily (top) and hourly (bottom) smartmeter data as input.

Transformer Type N...	Underground			Overhead		
City	ACCURACY OF INTERRUPTION PREDICTION	Predicted Failure	Failed	ACCURACY OF INTERRUPTION PREDICTION	Predicted Failure	Failed
BATESBURG	79%	2,071	1,705	98%	4,019	4,088
CHAPIN	74%	67,671	53,496	98%	134,369	132,010
GILBERT	.	17	0	.	27	0
IRMO	66%	175,033	130,339	95%	373,422	356,338
LITTLE MOUNTAIN	35%	5,044	3,062	98%	20,187	19,772
PELION	89%	24,736	22,367	99%	40,277	40,579
PROSPERITY	.	434	0	97%	2,372	2,314

Display 2. Transformer failure forecasts using weather and consumption forecasts as input.

.

CONCLUSION

Analytics performance capabilities on Intel® platforms and SAS® 9.4 Analytics software allows us to process full granularity data sets without compromising data quality. Working with such full scale data allows utility companies to be much more precise in forecasting, with a direct impact on short-term and long-term operation costs. Having and using all this data allows us to perform experimental ad hoc analysis, to derive impactful business insights.

Furthermore, by incorporating analysis at the edge, in the moment that events happen, we can predict potential equipment failures, resulting in reduced maintenance costs and operating costs.

REFERENCES

"Intel IoT Gateways." . 7/31/2015. Available at https://www-ssl.intel.com/content/www/us/en/embedded/solutions/iot-gateway/overview.html.

"Intel® Xeon®.Processor E5 Family" . 7/31/2015. Available at https://www-ssl.intel.com/content/www/us/en/processors/xeon/xeon-processor-e5-family.html.

"Intel® Xeon®.Processor E5 Family" . 7/31/2015. Available at https://www-ssl.intel.com/content/www/us/en/processors/xeon/xeon-processor-e7-family.html.

"How Does Forecasting Enhance Smart Grid Benefits?" . 7/31/2015. Available at http://www.sas.com/en_us/whitepapers/how-does-forecasting-enhance-smart-grid-benefits-106395.html.

"When One Size No Longer Fits All - Electric Load Forecasting with a Geographic Hierarchy" . 7/31/2015. Available at http://www.sas.com/en_us/whitepapers/when-one-size-no-longer-fits-all-106226.html

"SAS Event Stream Processing" . 7/31/2015. Available at http://www.sas.com/en_us/software/data-management/event-stream-processing.html

"SAS Visual Analytics", 7/31/2015, Available at http://www.sas.com/en_us/software/business-intelligence/visual-analytics.html.

ACKNOWLEDGMENTS

We would like to acknowledge Lorry Hardt at SAS® Institute with his help with devising the consumption forecasting and transformer failure prediction models and gathering the actual forecast results.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Suleyman Sair
Enterprise: Intel Corporation
Address: 5000 W. Chandler Blvd. Mail Code CH7-401
City, State ZIP: Chandler, AZ, 85226
Work Phone: 480-554-1357
E-mail: Suleyman.Sair@intel.com