modius®

REMOVE COMPLEXITY | EXTRACT MORE VALURE FROM IOT DATA

IoT in a Box

Removing Complexity in IoT deployments

Many organizations are striving to define, rationalize and implement IoT solutions for their business units and customers. While the promise of IoT is great – the obstacles for a successful implementation can be daunting. This white paper will address some of the major challenges with implementing an IoT solution, and how some partners of Intel are breaking ground to reduce these impediments to launching IoT programs.

Understand IoT data well enough to unlock actionable intelligence can be a challenge and correlating those changes over time to understand the larger economic picture of your intelligent systems can seem almost impossible.

While the business case for IoT is often discussed, the ROI for an IoT implementation is highly dependent on the business value of the resulting system and the availability of pre-existing instrumentation for those “things”.

An overarching challenge is how to get started on an IoT project, and how to analyze all of the data that will result from a typical IoT implementation. Working together, two of Intel’s IoT Partners Dell EMC and Modius can now provide a proven approach for the integration, data management and analyses of IoT data with a pre-packaged end-to-end IoT solution.

These IoT Bundles are literally an “IoT in a Box” approach to getting IoT projects started easily and reliably

The IoT Data Challenge:

How to collect, manage and analyze IoT data

IoT devices can provide a wealth of data regarding equipment operational status and efficiency metrics. Once collected, this device data is available for a range of analysis from simple deterministic event definition to more powerful machine learning methods that are now ubiquitous.

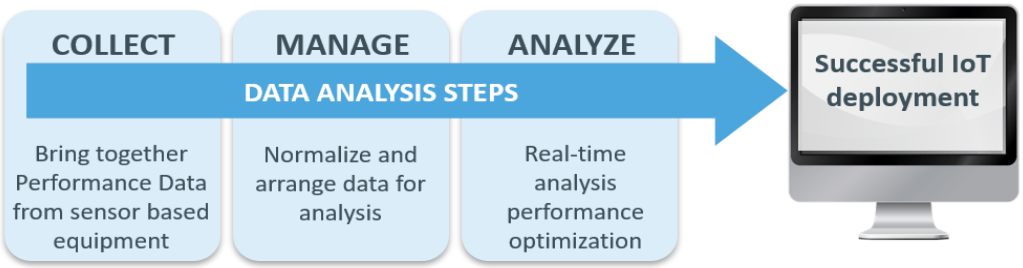
The type of analysis that can be performed is limited only by someone's imagination. However, before any analysis can take place, the data must be properly prepared.

Preparation Includes:

1. Handling missing values
2. Scaling and correcting data
3. Eliminating inconsistently recorded dates and times
4. Detecting and resolving data entry errors
5. Feature Engineering for machine language optimization
6. Adjusting the data to fit a normal distribution when necessary

It is estimated that up to 80%[1] of the effort in an IoT data analysis project is spent simply preparing the data for analysis. That estimate also assumes that some of the data has been collected and recorded by hand. It seems logical that when we are collecting data from smart IoT devices the data should be much cleaner and require far less cleaning. But, is that always a valid assumption?

In this paper we will address these challenges when implementing an IoT Solution, and how the Dell - Modius OpenData Bundles address these three challenges:



Data Collection

All critical facilities contain equipment and sensors made by several different vendors that have different useful lifetimes. So, there are some items from Vendor A, some from Vendor B, some that are old and some that are new, some that are critical to the entire facility (e.g. UPS, chillers, etc.) and some that support specific applications (e.g. NC equipment, conveyor belts, servers, temperature sensors, etc.) Most, if not all, of these devices are capable of being monitored using OT protocols like BACnet, Modbus, HART, SCAN, SNMP, etc. This device diversity presents several challenges for eventual upstream analysis of this IoT data:

First, with different devices from different vendors there is likely to be several different representations for similar data (this condition may also be true for different devices from the same vendor). This leads to a cacophony of data values. Sometimes voltages are recorded as integers in units of millivolts while other times being recorded as floating-point numbers in units of volts. Sometimes timestamps are represented as character strings and sometimes as an integer representing the number of seconds since Jan 1, 1970. Even though data is being collected and stored programmatically, there is potentially a lot of inconsistency in the data. The effort required to prepare this data for analysis can be significant and requires device specific knowledge for all devices.

Second, the data from devices is often accessed over legacy OT protocols such as BACnet, Modbus, SCAN, HART, SNMP etc. which are not secure and consume significant network (or cellular) bandwidth. Transporting machine data across an enterprise network can be resource intensive. Depending on network access and the native protocol of the device, collecting and transporting in native protocols can create significant security challenges. Polling frequencies will vary based on the sensitivity of the equipment and the likelihood that measurements will change significantly over a time interval.

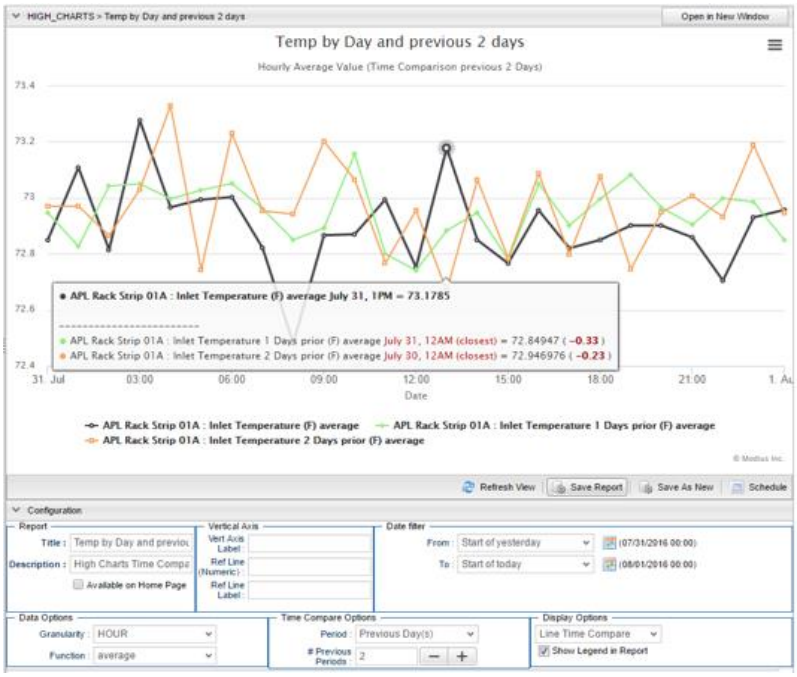
Data Management

The data collected from IoT devices is almost always time series data and meaningful analysis requires time synchronization. Since every device has its own clock this can be a problem. Analyzing data across multiple facilities in multiple time zones exacerbates the issue. Time synchronization of data is a prerequisite for all forms of time series analysis. Cleaning data that is not time synchronized can be difficult and time consuming at best and nearly impossible at worst.

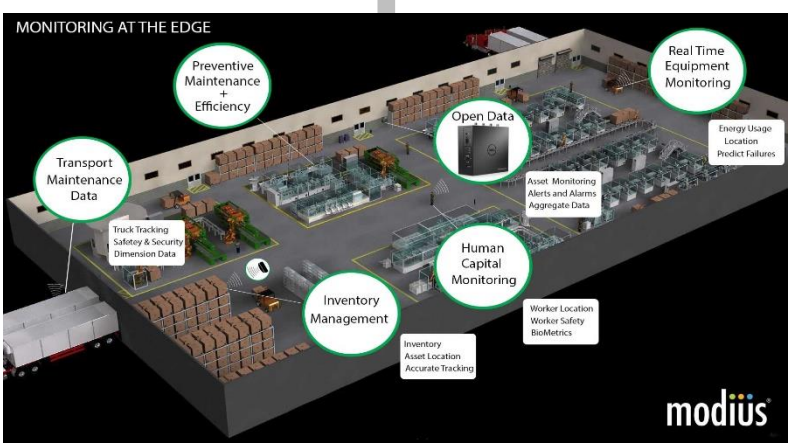
Additionally, statistically summarized aggregates of the time series data by hour, day, week and month allows different intervals to be efficiently chosen for upstream analysis without need for time-consuming data organization at run time. Furthermore, the enormous volume of raw sensor data can be rolled off as the relevance of the measurement becomes less sensitive over time. Minute by minute data will have less importance a year after the measurement when only general trends are of interest.

Data Analysis

In many cases, the devices and sensors in a critical facility were purchased and installed without any consideration for the eventual data analysis needed to achieve the business objectives of the organization. As a result, it may be necessary to do some feature engineering before starting the data analysis. Feature engineering applies knowledge of the physical site to create features that improve the analytics results. In some cases, it may be possible to reduce or eliminate the need for feature engineering by using computed points. Computed points are a mechanism for creating new features by applying mathematical calculations to existing data. These computed points need to be time-synchronized with the underlying raw point data to ensure consistency between the computed points and the underlying measured values.



Once collected, the data must be easily accessible for visualization and data mining. End users need to have easy to operate query access for Real time Operational Intelligence (RtOI). Data patterns and consistent reporting must be integrated with the data so analyses can be conducted ad hoc.



Conclusion

There are many challenges associated with the design and implementation of an end to end IoT solution. Data collection, data management and data analysis each have specific technical and functional complexities to consider during the development and deployment process. Modius OpenData has been designed to resolve these many data issues by processing machine data “at the Edge”, so data normalization and time synchronization can be accomplished at “first touch” and data transport can be accomplished in conformance with current IT standards. OpenData uses predefined templates for connecting to each device. These templates dramatically reduce the amount of time (and technical expertise) required to configure OpenData for a specific facility or type of equipment. Additionally, these templates ensure that every data point on every device is collected and scaled consistently. This consistency ensures that data can by confidently analyzed without the need for upstream "cleaning" or scaling. The following table provides a simple example of Modius’ encoding for IoT data values:

Data element	Type	Units
Voltage	Floating point	Volts
Temperature	floating point	Degrees F
Timestamp	Integer	Seconds since Jan 1, 1970

When using OpenData, all data points are time synchronized - even when collecting data in different time zones. Time synchronization is important for two reasons: First, time synchronized data can be used for diagnostics by providing a reliable order of events. Second, time synchronization dramatically reduces or eliminates the data preparation effort required for time series analysis. OpenData provides access to all the raw data for processing by third-party analytics tools but integrated analysis and visualization tools can often provide substantially complete analysis.

If the IoT implementation is supporting a Machine Learning project, feature engineering can be critical. Using OpenData, the process of collecting data and extracting it into tidy data frames can be performed without any machine learning skills. Items 1-4 on the list of data preparation tasks listed above can be completely resolved by using OpenData. Items 5 and 6 require some specific machine learning skills and data center knowledge.

By combining the Modius OpenData software with Dell Servers and IoT Gateways, enterprises can now purchase, install and configure an end-to-end IoT solution starting with a single SKU and finishing in 10 easy steps! These IoT Bundles provide the Dell servers, OpenData software, Dell IoT gateways and sample sensors so a production-ready IoT solution can be created in a matter of a few hours – and then extended gracefully to cover more devices at more sites once the business case and ROI can be finalized. These Bundles come in three sizes so an organization can pilot an IoT deployment that is representative of their production environment without complex pre-project analysis (which is often a guess at best) and the inevitable project delays caused by trying to discover all relevant device information. As the adage goes --- "Perfect is the Enemy of Good". By implementing a representative working IoT system, an organization can quickly validate cost assumptions and begin the journey towards the "perfect" IoT solution which is right-sized for their business. These IoT Bundles are designed to operate on-premise over enterprise IT infrastructure, but cloud-options are also available.

CONTACT YOUR MODIUS REPRESENTATIVE FOR MORE INFORMATION ABOUT HOW OPENDATA MAKES IT POSSIBLE TO EASILY DEPLOY IoT BUNDLES TODAY!

71 Stevenson Street, Suite 400, San Francisco, CA 94105 USA
sales@modius.com www.modius.com (888) 323-0066