PI World 2019 Lab

Taking your Data Science Projects to the Next Level with PI Data

OSIsoft, LLC
1600 Alvarado Street
San Leandro, CA 94577 USA
Tel: (01) 510-297-5800
Web: http://www.osisoft.com

Published: March 22, 2019

Table of Contents

Contents

1. Introduction

1.1 Learning Objectives

In this lab we will explore different tools for interacting with PI data as a data scientist. The lab will empower users with some data science background to better leverage data that their customers store in the PI System. You will learn how to explore PI data within the PI System before exporting it to develop a model. At the end of this lab, you will have gone through a small data science problem from data exploration to deployment using data stored in the PI System.

The assignment in the lab is to reduce the energy used to cool rooms at OSIsoft headquarters. Specifically, we will optimize the daily startup of the individual cooling units.

Every day, the VAVCO cooling units adapt to changing temperature, relative humidity, thermostat setpoints, building occupancy level, and other factors. The control system adjusts several factors to provide the necessary cooling to the rooms to ensure tenant comfort.

A pre-established occupancy schedule requires that rooms should be at a comfortable temperature between **7 AM and 7 PM** when employees are in the building. During those hours, the individual units bring the temperature to the desired setpoint. The units follow this schedule to meet that requirement:

- Turn on at some point in the morning and cool the room to setpoint by 7 AM
- Maintain room temperature at setpoint during occupied hours
- Shut-down at 7 PM when the building becomes unoccupied

The initial startup should finish as close to 7 AM as possible. Reaching the setpoint too early results in wasted energy cooling an unoccupied room. Conversely, if the setpoint is not reached by 7 AM, employees will not be comfortable in rooms that have not been cooled. The business unit believes that the current startup schedule could be improved by examining the historical data.

In other words, our objective is to predict how long a unit will take to reach the setpoint depending on current conditions to ensure the unit reaches the setpoint as close to 7 AM as possible.

1.2 Tasks

The aim of the lab is to demonstrate the PI System tools that enable different stages of a data science project. These steps follow the cross-industry process for data mining (CRISP-DM) framework. The tools covered in the lab allow data scientists to carry out important tasks in the project:

- Data understanding/exploration
- Data preparation/export
- Modeling and evaluation
- Deployment

1.3 Breakdown of Lab Content

Your objectives in this section are:

Data understanding/exploration. We leverage PI Client tools, including PI System Explorer, Event Frames, and PI Vision, to develop understanding about how the data has been structured as well as important trends.

Data preparation/export. We shape raw PI data into easily-consumed data tables using PI Developer Technologies, such as PI Web API and PI SQL Client, as well as the PI Integrator for Business Analytics. We discuss pros/cons of these different approaches.

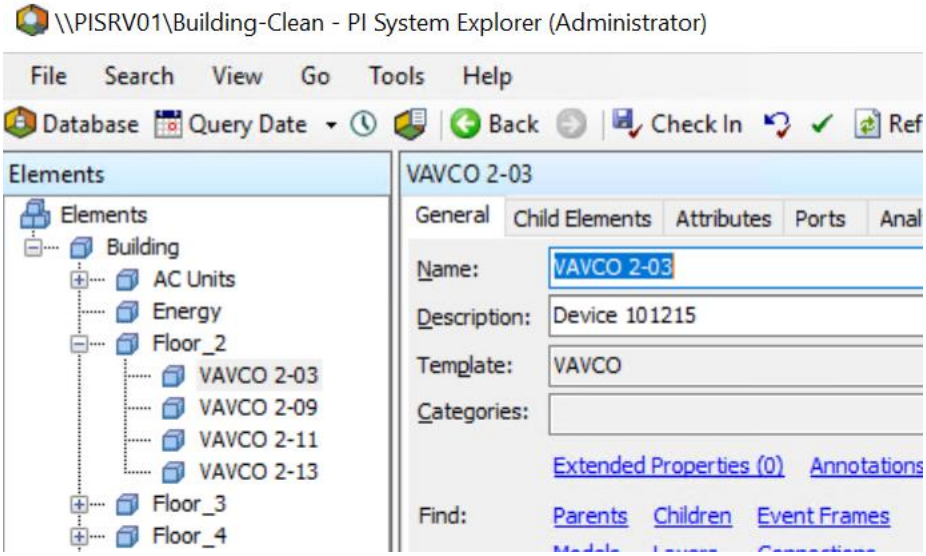
Modeling and evaluation. Using the data exported earlier in the lab, we develop a predictive model for the business problem in a Jupyter notebook. We then evaluate the model and export it in a format to allow deployment.

Deployment. With model in-hand, we review two deployment scenarios: a PI Web API-based approach and streaming from the PI Integrator for Business Analytics to an Apache Kafka server. The PI Web API-based approach includes writing data back to the PI System, while the streaming only covers data flow up to Kafka.

2. Directed Activity – Understand asset hierarchy in PI AF

2.1 Objective

We need to know where to find the data before beginning to develop a model. Unlike the PI administrators and subject matter experts (SMEs) who know their system intimately, we don't know what data is available, where to find that data, and how that data is structured. The aim of this exercise is to find the elements used to organize the data streams and identify what might be applicable to our problem.



2.2 Tasks

- Explore the AF hierarchy
- Identify PI tags/AF attributes available for analysis
- Think about which attributes would be relevant for the current problem
- Examine time series aggregation using event frames

2.3 Solution

- **Step 1.** Open PI System Explorer and select the Lab Building Data database
- **Step 2.** Expand the Building and Floor_2 elements, then select the VAVCO 2-03 element – note the attributes for that element
- **Step 3.** Examine the attributes associated with the Weather element under Building
- **Step 4.** Create a list of attributes that may be relevant for this project

- **Step 5.** Select the VAVCO startup event frame template in PI System Explorer and examine the attributes
- **Step 6.** Open the event frame trigger from the Analysis tab on the VAVCO template and interpret the trigger logic

2.4 Discussion

- What attributes are relevant in the problem that we're trying to solve?

- What did you learn from the analysis?

- What are some use cases where you might use Event Frames in your system?

 Tip	All elements based on the **VAVCO** template have the same attributes.

 Best Practice	PI System Explorer does have an ad-hoc trending feature, but it lacks many of PI Vision's features and is not recommended.

 Tip	Event frames are one way to shape time series values into a data science-friendly format. Other methods include interpolation/aggregation at fixed times and calculation of virtual features.

3. Directed Activity – Visualize time behavior using PI Vision

3.1 Objective

Since we're not experts in facilities management, it's important to start building intuition about how these data behave. This added context can help inform decisions down the line as well as communication with SMEs. Visualizing evolution in time is necessary to understand the time series data stored in the PI System. In this activity, we'll plot data trends during times of interest.

3.2 Tasks

- Create a new display in PI Vision that trends the values
- Compare startups for different VAVCO units in PI Vision
- Compare event frames in PI Vision

3.3 Deliverables

- Display that shows trends for a specific element and time range
- Display that compares event frames for different days/elements

3.4 Solution

- **Step 1.** Open Google Chrome and navigate to the bookmark labeled PI Vision. Use the same credentials that were used to logon to the VM.
- **Step 2.** Create a new display
- **Step 3.** Type VAVCO 6-09 in the search bar, select the asset, and drag % cooling from the attribute pane onto the display to create a trend
- **Step 4.** Change the start/end times in the bottom-left and bottom-right of the display to 14-Sep-17 02:30:00 and 14-Sep-17 05:00:00.
- **Step 5.** Drag the Room Temperature and Occupied Setpoint attributes to a new trend just below the % cooling
- **Step 6.** Right-click on the new trend, select Format Trend, and change "Single or Multiple Scales" to "Show single scale"
- **Step 7.** Select VAVCO 6-06 from the Asset dropdown and compare trends
- **Step 8.** Select VAVCO 3-16 from the Asset dropdown and compare trends
- **Step 9.** Select VAVCO 6-09 from the Asset dropdown again
- **Step 10.** Click the Events tab on the PI Vision display, right click on the event frame, and select "Apply Time Range"
- **Step 11.** Right click on the event frame again, select "Compare Similar Events by Type" (save the display if prompted), and examine the overlaid trends

3.5 Discussion

- What have we learned by visualizing the data?

- How would visualizing a handful of random cooling events help with analyzing thousands of data points?

- What other resources can we use to better understand a data set?

4. Directed Activity – Explore data using PI Web API libraries

4.1 Objective

Open source libraries have become increasingly popular for data science, particularly those based on Python and R. These tools can access PI System data via the PI Web API, a REST endpoint for the PI System. We've created sample code to facilitate data access for the lab using a Jupyter notebook.

Documentation is available for building the API calls from scratch, but for the purposes of the lab we'll use some pre-written code to simplify the data access. Building a PI Web API library for your production system is beyond the scope of this lab, though interested students can review the material covered in the "Build PI Applications Faster with PI Web API Client Libraries" talk by Marcos Loeff.

In [5]:
```
from WebAPIHelper import *
import matplotlib.pyplot as plt
import matplotlib.dates as md

webapiurl = 'https://pisrv01/piwebapi'
dataarchive = 'PISRV01'
afserver = 'PISRV01'
afdatabase = 'Lab Building Data'
client = PIClient(webapiurl,dataarchive,afserver,afdatabase)
```

4.2 Tasks

- Load event frame data into a Jupyter notebook
- Cleanse event frame data
- Find attributes that correlate with event frame duration

4.3 Deliverables

- Jupyter notebook that includes cleansed event frame data
- Regression plots that indicate the relationship between variables

4.4 Solution

- **Step 1.** Launch Jupyter Notebook and open “Data exploration – PI Web API”
- **Step 2.** Follow the instructions in the notebook to connect to the PI System, load event frame data, and examine the first few rows of data
- **Step 3.** Follow the instructions in the notebook to clean the raw data such that attributes have correct data types and missing values handled appropriately
- **Step 4.** Filter event frames that reach the setpoint and examine correlation between key factors
- **Step 5.** Guess which factors most important for predicting time to cool a room

4.5 Discussion

- What was the most difficult part about bringing this data into the Jupyter notebook?

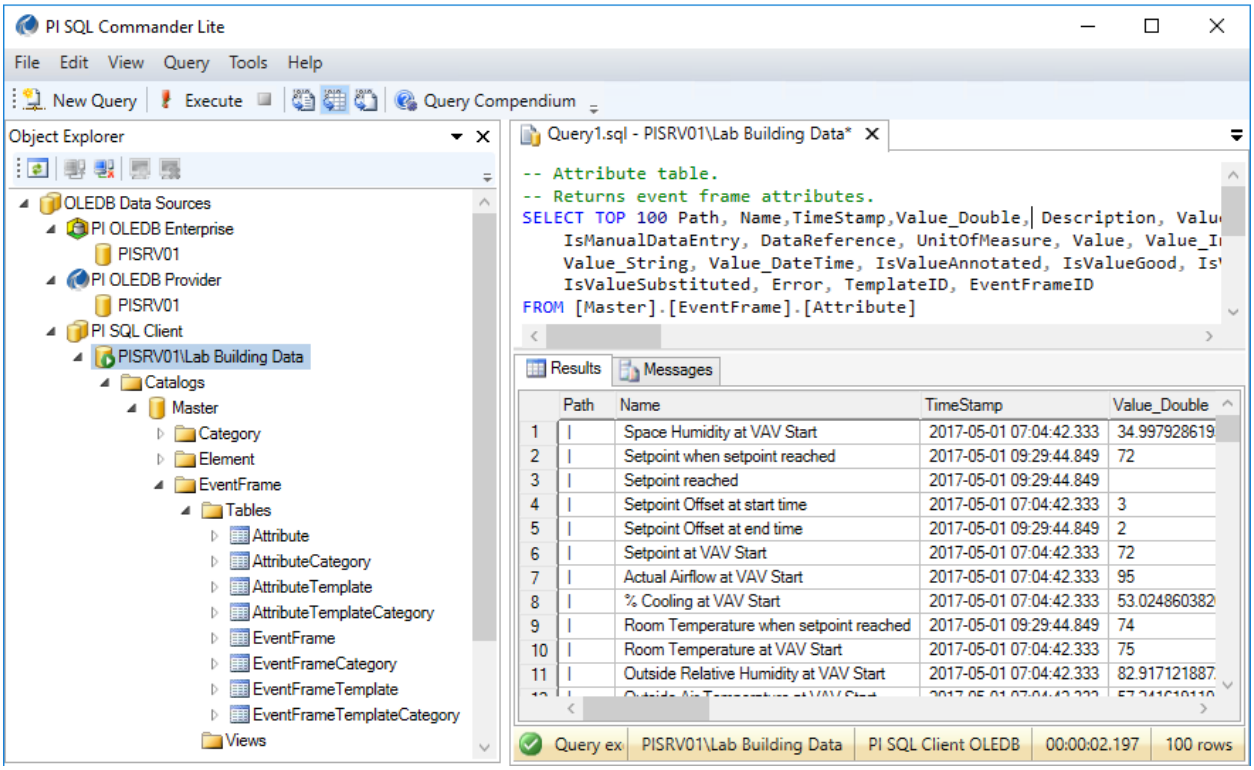
- What limitations might there be to using this approach for data access?

 Tip	To avoid repeatedly querying the PI System for data, it can be helpful to split the code into multiple parts: the first queries the PI System, cleanses the data, and writes the result to disk, while the second accessed the cleansed data for exploration. This minimizes unnecessary queries against the PI System.

5. Directed Activity – Explore data using PI RTQP Engine

5.1 Objective

If PI Web API is unfamiliar to your team, or you're most comfortable using SQL syntax, then the PI SQL Client/RTQP Engine may be a preferable way to access PI data. This approach uses SQL-like queries to shape and load PI data. Pre-defined sample queries show how to access the data. The returned values are further processed in the client.



5.2 Tasks

- Query event frame data using PI SQL Client pre-defined queries
- Access the data from a Jupyter notebook using modified query
- Pivot the data to a more usable format

5.3 Deliverables

- Sample query in PI SQL Commander that returns data for the project
- Jupyter notebook that includes a similar dataframe to the previous exercise, but with the data obtained using the PI SQL Client

5.4 Solution

- **Step 1.** Open PI SQL Commander Lite (64-bit) from the start menu
- **Step 2.** Right click on “PISRV01\Lab Building Data” under “PI SQL Client”, select “Connect”, then click “OK”
- **Step 3.** Find the Event Frame Attribute table and execute the predefined query
- **Step 4.** Look at the query results and the query itself. Identify the changes that you might want to make in the query before bringing it into the client.
- **Step 5.** Open the “Data exploration – PI SQL Client” Jupyter notebook and run the query using the “win32com.client” module.
- **Step 6.** Reshape the data to a form that can be used in further analyses.

5.5 Discussion

- What are the advantages of this tool over the PI Web API? What are the additional limitations?

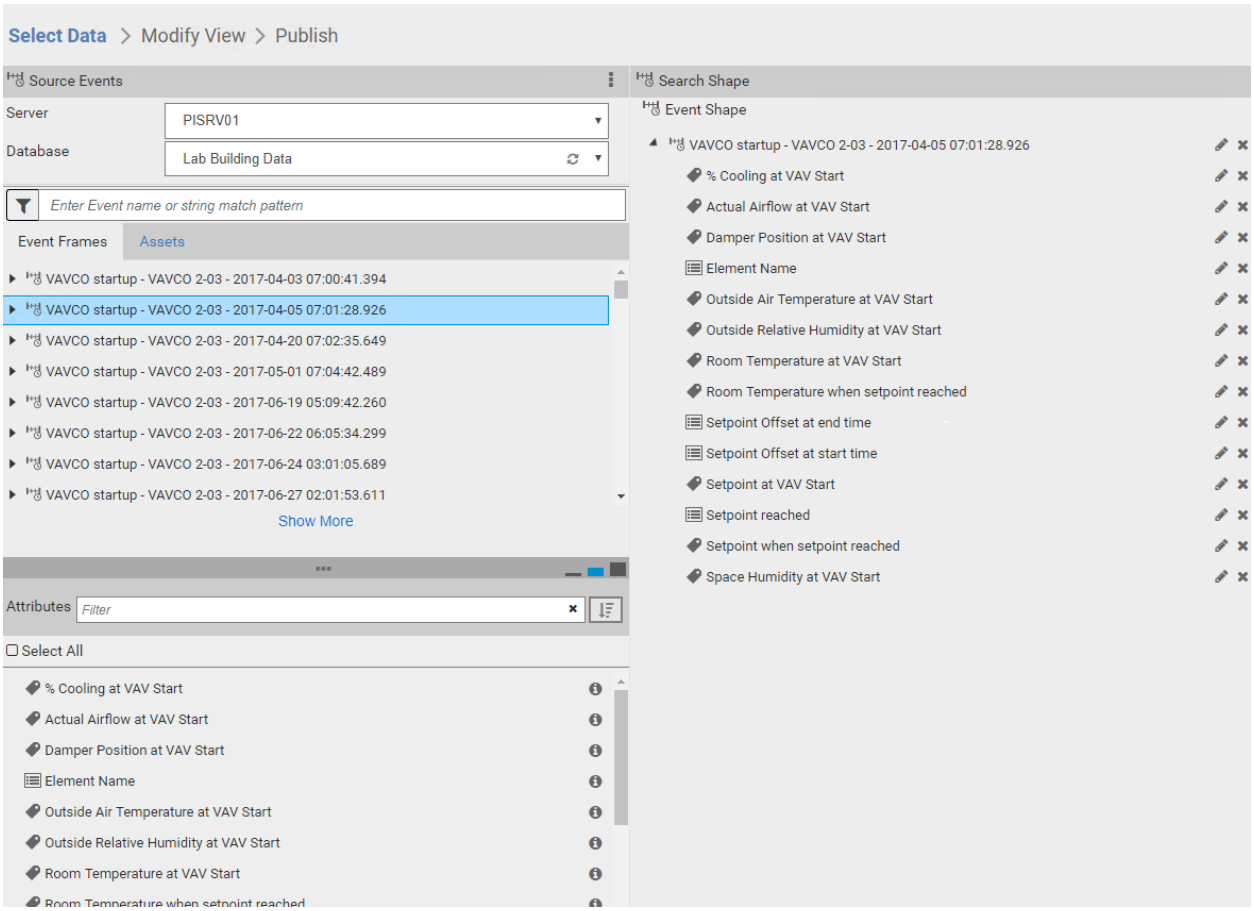
- Which data access technology did you prefer?

- Would users in your organization be more likely to use PI Web API (REST) or PI SQL Client? Why?

6. Directed Activity – Export data using the PI Integrator for Business Analytics

6.1 Objective

We'll use Python to build a model to describe the data. The model requires an input of historical data for fitting and validation. Rather than accessing the PI data each time we run the model, the PI Integrator for Business Analytics will automatically shape and export the PI data to a flat file that can be read for analysis.



6.2 Tasks

- Shape data as an event view using the PI Integrator for Business Analytics
- Publish data table to a CSV file

6.3 Deliverables

- Text file with a year of Event Frame data in CSV format

6.4 Solution

- **Step 1.** Open Google Chrome and navigate to https://pisrv01.pischool.int:4444
- **Step 2.** Create an Event View that includes attributes of interest in the Search Shape
- **Step 3.** Review the data preview and adjust the data format
- **Step 4.** Publish the view to a Text Output

6.5 Integrator Tips

- Filter for the VAVCO startup Event Frame Template rather than the Event Frame Name
- Change the Duration column in the preview to show minutes rather than hours
- Filter out startup periods where the setpoint was not reached
- Include event frames between 01-Mar-2017 and 01-Feb-2018

6.6 Discussion

- Open the text file using Notepad or Excel. Does it contain the expected data? How is it different?

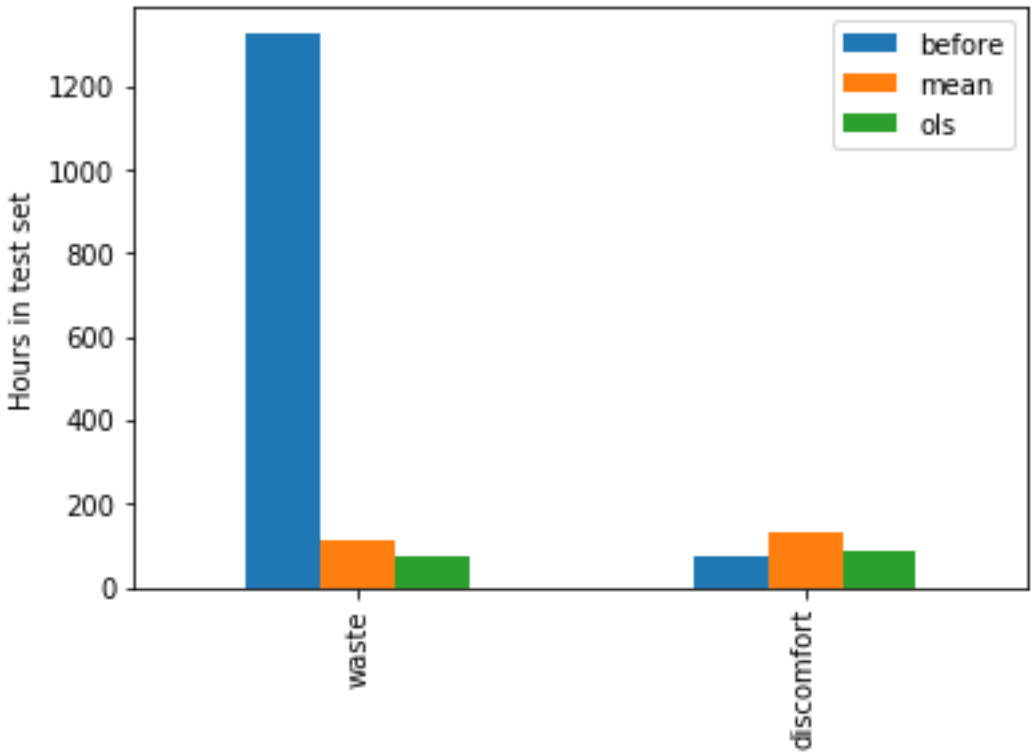
- Consider the differences between the integrator and PI Developer Tools. When might you use one over the other?

✓ Best Practice	The PI Integrator for Business Analytics is best used to combine many AF attribute and/or event frame values in a structured data table. Data science applications read from these quickly without affecting operation of the PI System.

7. Directed Activity – Train and evaluate a model

7.1 Objective

The purpose of this activity is to develop a model based on the data exported in the previous activity using a Jupyter notebook. Data access is much simpler as standard libraries can read from CSV files without difficulty.



7.2 Tasks

- Connect to the table published by the PI Integrator for Business Analytics using a Jupyter notebook
- Fit a model to predict cooling duration using standard data science libraries

7.3 Deliverables

- Model that predicts the time to cool a room based on current conditions
- Validation that the model is more accurate than a naïve solution

7.4 Solution

- **Step 1.** Open the Jupyter notebook titled “Model Training and Evaluation”
- **Step 2.** Follow the instructions in the notebook
- **Step 3.** Evaluate model performance using bar charts

7.5 Discussion

- What other models could we have used for this problem? What tradeoffs are involved?

- What other metrics can we use to evaluate the model performance?

- How would you have approached this problem differently?

✓ Best Practice	A properly-designed model should perform better than both historical averages and a naïve solution. A “naïve” solution can be a simple average of past values for numerical data, the previously-measured value for time series, or the most common value for categorical data.

8. Directed Activity – Deploy model using PI Web API

8.1 Objective

The purpose of this activity is to run the model periodically to calculate new estimates for the time to cool the room then write those values back to PI. This activity uses the PI Web API to read PI System data and to write the forecasts back to the PI System.

```
Time to reach setpoint
Time: Wed Aug 29 08:09:15 2018
VAVCO 2-03: 25.1 minutes
VAVCO 2-09: 23.3 minutes
VAVCO 2-11: 23.6 minutes
VAVCO 2-13: 24.0 minutes
VAVCO 3-09: 24.4 minutes
VAVCO 3-10: 25.1 minutes
VAVCO 3-11: 1.9 minutes
VAVCO 3-15: 24.0 minutes
VAVCO 3-16: 23.6 minutes
VAVCO 4-03: 3.0 minutes
VAVCO 4-15: 24.4 minutes
VAVCO 4-16: 25.1 minutes
VAVCO 4-17: 25.5 minutes
VAVCO 5-09: 24.4 minutes
VAVCO 5-10: 25.1 minutes
VAVCO 5-11: 25.1 minutes
VAVCO 5-12: 24.8 minutes
VAVCO 5-13: 3.0 minutes
VAVCO 6-06: 3.8 minutes
VAVCO 6-07: 1.9 minutes
VAVCO 6-08: 1.5 minutes
VAVCO 6-09: 2.7 minutes
VAVCO 6-10: 3.0 minutes
VAVCO 6-11: 2.7 minutes
```

8.2 Tasks

- Load the model developed in the previous section
- Read recent values from the PI System as inputs to generate a prediction
- Write the prediction back to the PI System

8.3 Deliverables

- Console that indicates the predictions for each unit
- New values written to AF attributes

8.4 Solution

- **Step 1.** Review the file Deployment_Streaming.py
- **Step 2.** Run the file with the command “python Deployment_Streaming password”, where password is the pischool\student01 password

8.5 Discussion

- What are the limitations of using this approach to deploy the models?

- What other options are available for model deployment?

9. Directed Activity – Publish data to Kafka using the PI Integrator for Business Analytics

9.1 Objective

If you don't want to query the PI System using PI Web API but still want to stream values as they arrive in the PI System, the PI Integrator for Business Analytics can stream values to a Kafka server. Apache Kafka has some advantages over a standalone service and the PI Integrator automatically handles a wide range of situations. The integrator can also stream to other cloud targets in Azure and AWS.

9.2 Tasks

- Start Zookeeper and Kafka server
- Configure the integrator to stream to the server
- Check that the server is receiving new values

9.3 Deliverables

- Command prompt confirms JSON object that contains recent values

9.4 Solution

- **Step 1.** Open the command prompt and run C:\Lab\Kafka\start_zookeeper.bat
- **Step 2.** Open the command prompt and run C:\Lab\Kafka\start_kafka.bat
- **Step 3.** Configure a streaming view with the PI Integrator for Business Analytics
- **Step 4**. Check the topic publication from the command prompt (replace `TopicName` with the name of the streaming Integrator view): `C:\Lab\Kafka\bin\windows\kafka-console-consumer.bat --bootstrap-server localhost:9092 --topic TopicName --from-beginning`

9.5 Common Issues

- The Kafka server occasionally has difficulty starting up. If you see a fatal error in the console, delete the files in C:\tmp\kafka-logs and try again.

9.6 Discussion

- What are the potential advantages of streaming with the integrator compared with other PI Developer Technologies?

- How did creating the streaming view compare to the PI Web API code?


Have an idea how to improve our products?
OSIsoft wants to hear from you!
https://feedback.osisoft.com/

Save the Date!

OSIsoft PI World Users Conference in Gothenburg, Sweden. September 16-19, 2019.

Register your interest now to receive updates and notification early bird registration opening.

https://pages.osisoft.com/UC-EMEA-Q3-19-PIWorldGBG-RegisterYourInterest_RegisterYourInterest-LP.html?_ga=2.20661553.86037572.1539782043-591736536.1533567354