

# PI World 2019 Lab

PI Integrator for Business Analytics: In-depth Tutorial



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

© 2019 by OSIsoft, LLC. All rights reserved.

OSIsoft, the OSIsoft logo and logotype, Analytics, PI ProcessBook, PI DataLink, ProcessPoint, Asset Framework (AF), IT Monitor, MCN Health Monitor, PI System, PI ActiveView, PI ACE, PI AlarmView, PI BatchView, PI Vision, PI Data Services, Event Frames, PI Manual Logger, PI ProfileView, PI WebParts, ProTRAQ, RLINK, RtAnalytics, RtBaseline, RtPortal, RtPM, RtReports and RtWebParts are all trademarks of OSIsoft, LLC. All other trademarks or trade names used herein are the property of their respective owners.

#### U.S. GOVERNMENT RIGHTS

Use, duplication or disclosure by the U.S. Government is subject to restrictions set forth in the OSIsoft, LLC license agreement and as provided in DFARS 227.7202, DFARS 252.227-7013, FAR 12.212, FAR 52.227, as applicable. OSIsoft, LLC.

Published: April 18, 2019

# Table of Contents

## Contents

Table of Contents .....	3
1. Introduction.....	5
1.1 Overview of Lab .....	5
2. Directed Activity – What is the CRISP-DM methodology?.....	6
2.1 Objective of Activity .....	6
2.2 Tasks .....	6
2.3 Business Understanding and Data Understanding .....	6
2.3.1 Business Understanding.....	6
2.3.2 Data Understanding .....	7
2.4 Data preparation.....	7
2.5 Modeling, Evaluation and Deployment .....	8
2.5.1 Modeling.....	8
2.5.2 Evaluation .....	8
2.5.3 Deployment.....	8
3. Directed Activity – Turbine efficiency - Heat Recovery Steam Generators.....	9
3.1 Objective of Activity .....	9
3.2 Tasks .....	9
3.3 Business understanding .....	9
3.3.1 Scenario.....	9
3.3.2 Produce a project plan.....	10
3.3.3 Identify business success criteria .....	10
3.4 Data understanding .....	10
3.4.1 Data Exploration.....	10
3.5 Data Preparation .....	12
3.5.1 Define a new target .....	12
3.5.2 Create an asset view.....	12
3.5.3 Create an asset shape .....	13
3.5.4 Modify the asset view's publication parameters .....	14
3.5.5 Monitoring the view .....	15
3.5.6 Viewing the data in a BI tool .....	16
4. Directed Activity – Well Drilling - Stick-slip events.....	17

---

4.1	Objective of Activity .....	17
4.2	Tasks .....	17
4.3	Business understanding .....	17
4.3.1	Scenario .....	17
4.3.2	Produce a project plan.....	18
4.3.3	Identify business success criteria .....	18
4.4	Data understanding .....	18
4.4.1	Data Exploration.....	18
4.5	Data Preparation .....	19
4.5.1	Create a summarized event view .....	19
4.5.2	Modify the Event View's publication parameters.....	21
4.5.3	Monitoring the view .....	22
4.5.4	Viewing the data in a BI tool .....	22
4.5.5	Create a dataset with granular data.....	23
4.6	Modelling, Evaluation and deployment .....	23
4.6.1	Modelling.....	23
4.6.2	Evaluation, deployment .....	24
Appendix A	.....	25
Sampling interval – Basics .....		25
Exception reporting.....		25
Compression testing:.....		25
Scan class and offset .....		26
Non-uniform data.....		27
Appendix B	.....	28
Five parameters are relevant when configuring a Continuous View. ....		28
Save the Date!.....		30

# 1. Introduction

## 1.1 Overview of Lab

The **PI Integrator for Business Analytics (BA)** is a data science enablement tool. It exports PI System data in a consumable format for **Business Intelligence (BI)** tools and analytics engines. As the user in charge of creating the views and administering the PI Integrator for Business Analytics, there are important concepts to understand both on the PI System side and on the data science side. Students will learn about **CRISP-DM** - a common methodology used for data mining, how to prepare PI System data and to ensure that the PI Asset views and PI Event views created in the PI Integrator for BA will contain the relevant data that they need.

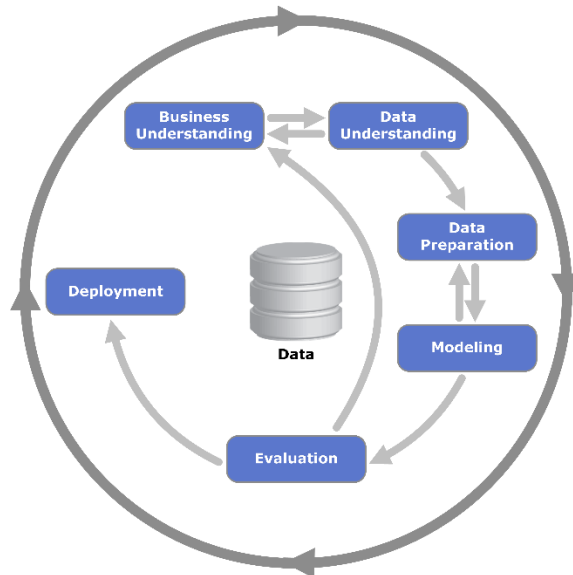
## 2. Directed Activity – What is the CRISP-DM methodology?

### 2.1 Objective of Activity

In this section we review the steps involved in the CRISP-DM methodology and see in which steps the PI Integrator for Business Analytics fits into the methodology.

### 2.2 Tasks

- Understand the steps involved in the CRISP-DM methodology



In this course, we will focus understanding of the first three phases: the business understanding, data understanding and data preparation steps.

### 2.3 Business Understanding and Data Understanding

The business and data understanding steps are closely dependent on each other and often goes back and forth.

#### 2.3.1 Business Understanding

Your first objective is to understand, from a business perspective, what results you want to achieve. Your organization may have competing objectives and constraints that must be properly balanced. The goal of this stage of the process is to uncover important factors that could influence the outcome of the project. Neglecting this step can mean that a great deal of effort is put into producing the right answers to the wrong questions. (<https://www.sv-europe.com/crisp-dm-methodology/>) It is

recommended to involve as many people as necessary to get a solid understanding of the business problem that we want to resolve.

1. **Set objectives:** What is the primary objective? Document specific business objectives decided upon by key decision makers.
2. **Produce a project plan:** What is the plan for achieving this data mining and business goals? What tools will we be using? In this workbook, we make the assumption that the PI Integrator for Business Analytics has been selected as one of the tools to be used during the data preparation step.
3. **Business success criteria:** Define as precisely as possible the criteria to determine whether the project was successful.

**Relevance to the PI System:** What data is available in PI can help us achieve this goal? Are the relevant KPIs already calculated in PI so that progress can be monitored?

### 2.3.2 Data Understanding


The next step is to acquire the data required for the project as defined in the Business Understanding step. Inspect what data is available and understand the relationship between data sources. Does additional data need to be acquired? Out of all the data available, what variable seems to be the most promising?

**Relevance in the PI System:** What data streams are already available in PI? What data do we need to combine it with, to produce meaningful results? How are the PI tags configured? Can we work with interpolated data?

**TIP:** If exception and compression settings are set up correctly in the PI System, then interpolated data can be trusted. See appendix A for more details.

## 2.4 Data preparation

In this step, you will select your data, cleanse your data, construct the required data and integrate all the data. It is estimated that 50%-70% of the time spent on a data science project is spent on this step. This is the step that the PI Integrator is designed to facilitate with respect to the OT data contained in the PI System. The PI Integrator allows you to select, clean and construct data from the PI System and publish it to a target destination, where it can be retrieved during the next step of the process, the modeling step.

 Tip	Verify the requirements of the modeling tools that you plan to use!
--	---

## 2.5 Modeling, Evaluation and Deployment

### 2.5.1 Modeling

Using the data prepared, we can now use the analysis tools selected during the business understanding phase to generate a model for the data. This is an iterative process with the data preparation step as you may find that you will need to go back and prepare the data in a different manner.

In this step, we also need to ensure that the model built is statistically accurate, valid and satisfies the business success criteria.

**Note:** The data in the AF example kits used in this lab is randomly generated. No predictive model will be created in this lab.

### 2.5.2 Evaluation

Once the model has been generated, we assess the results and determine if the model satisfies the business objectives. Is the model usable for future deployment? Is there anything that we've overlooked during the entire process?

### 2.5.3 Deployment

In this final phase, we plan for the deployment, monitoring and maintenance, and a final report. Predictive analytics can help to improve the operations KPI, or patterns can be recognized and used for planning and marketing.



## 3. Directed Activity – Turbine efficiency - Heat Recovery Steam Generators

### 3.1 Objective of Activity

Examine a heat recovery steam generator system and follow the first 3 steps of the CRISP-DM methodology to generate a dataset for further analysis.

### 3.2 Tasks

- Perform the business understanding step of the CRISP-DM methodology
- Perform the data understanding step of the CRISP-DM methodology
- Perform the data preparation step of the CRISP-DM by creating an asset view

### 3.3 Business understanding

#### 3.3.1 Scenario

A combined cycle power plant generates electricity in two cycles. A gas turbine burns fuel and generates electricity and hot exhaust air. A Heat Recovery Steam Generator (HRSG) captures the exhaust heat, turns the energy into steam and delivers it to a steam turbine. The steam turbine in turns produces more electricity, increasing the overall efficiency and electrical output of the system.

In this example, the HRSG contains multiple phases. A high-pressure steam turbine leads the intermediate and low-pressure turbines. This multi-step process increases the overall efficiency of the system.

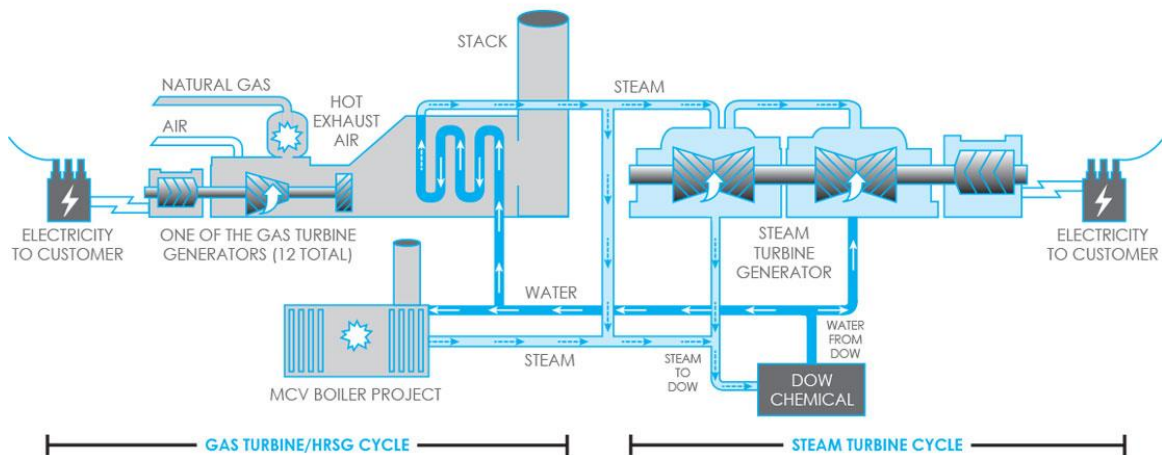


Figure 1: Combined-cycle power plant

In this scenario, we are concerned about the efficiency of the high-pressure recovery cycle of the HRSG. The turbine is rated for an efficiency of 86.5% but has been running around an average of 84.8%. A 1% increase in efficiency at this scale would result in an increase in revenue from electricity generated of roughly 400,000\$ per year.

**Discussion:** What business objective have you encountered in your own workplace with your operational data?

### 3.3.2 Produce a project plan

Who are the stakeholders in such a project? What kind of data can we look at?

The plant manager, the project manager, the process engineer, the operators.

- The project manager plans on sending the data and doing a preliminary analysis in a BI tool and creating a predictive model of the high-pressure steam pressure and temperature.
- The plant manager monitors the efficiency KPI on a PI ProcessBook dashboard. He declares that the high-pressure steam turbine 1 has never performed as well as the other turbine on site.
- The process engineer has told us that the data for the steam turbine resides in the PI System. The process engineer suspects that it is due to a problem with feed quality, and not the turbine itself. He says that the high-pressure steam temperature and pressure are not optimal. The high-pressure steam comes from the high-pressure superheater.
- The operators believe that the steam quality suffers because the boiler has difficulty retaining a constant water level.
- The data will be sent to the local SQL server where it can be analyzed by Power BI and accessed by our modeling tool.

### 3.3.3 Identify business success criteria

Based on our discussion so far, what could be the business success criteria?

Increased efficiency of the high-pressure cycle of the steam turbine by 1% in 2019

## 3.4 Data understanding

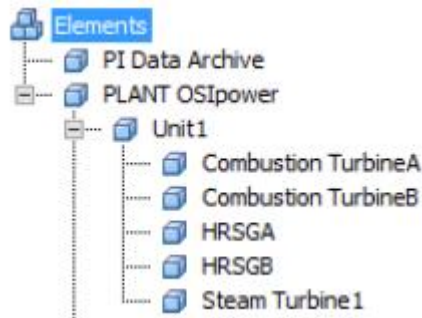
### 3.4.1 Data Exploration

Let's explore the data we have available to us in the AF Database. Find all the attributes that will be helpful for analysis.

Log on to the server PISRV01 and open PI System Explorer. Connect to the following AF Server and AF database:

**AF Server:** PISRV01

**AF Database:** OSIDemo Turbine Efficiency FULL



This plant has two units, Unit1 and Unit2. Each unit has two combustion turbines and two HRSGs. The HRSGs feed the steam turbine.

Take 5 minutes to explore the AF Hierarchy and find relevant attributes based on the requirements in the business understanding section above.


**Solution:** The attributes we are interested in are as follows:

- Cold Reheater Steam Pressure
- Cold Reheater Steam Temperature
- SH Steam Outlet Pressure
- SH Steam Outlet Temperature
- HP Economizer Feedwater Temp
- HP Feedwater Flow
- HP Drum level
- HP Drum pressure

After some analysis of the data, we determine the following criteria:

- PI tag values update once a minute, with a 10 second offset to the top of the minute. In other words, the first values of the day are timestamped 00:00:10, 00:01:10, 00:02:10, ...
- We will be sampling every 5 minutes, and we will work with the 5-minute average of the process tags for the high-pressure superheater, instead of the real- measurements. The high-pressure superheater attributes listed above all start with the prefix HP.
- Our data scientist will start with the data from 2018, as a first pass analysis.


- There is a pressure relief valve in the system. Any pressure readings over 2000 psia are instrument errors, and therefore should be filtered out.

 Tip	<p>Get familiar with your data.</p> <ul style="list-style-type: none"> <li>• If you are working with peaks, use event frames or filters.</li> <li>• For high frequency data that does not need to be sampled frequently, consider moving averages</li> <li>• In most cases, your data should be non-uniform due to exception and compression. Understand that you can still interpolate your data without risk if configuration is done correctly.</li> </ul> <p>See appendix A for more information.</p>
--	---

## 3.5 Data Preparation

### 3.5.1 Define a new target

In the project planning phase, we decided to send the data to the local SQL server. To define this output destination, we create a target in the PI Integrator for Business Analytics.

- 1) Connect to the PI Integrator UI using the Chrome shortcut on the desktop: <https://pisrv01>. There is also a bookmark available.
- 2) Go to the administration section under the pancake menu .
- 3) Under Targets, add a publish target of target type Microsoft SQL Server. An example target called SQL is pre-configured, so name yours **SQL\_<Student Name>**.
- 4) The hostname of the SQL server is **pisrv01**. Once connected, select **SQLDatabase** as the destination SQL database.
 

**Note:** SQL Authentication parameters are optional. By default, the service account identity is used.
- 5) Verify the SQL writer, and then save the changes.

### 3.5.2 Create an asset view

An asset view is used when we are interested in data related to PI AF assets, as opposed to event views, which are related to event frames.

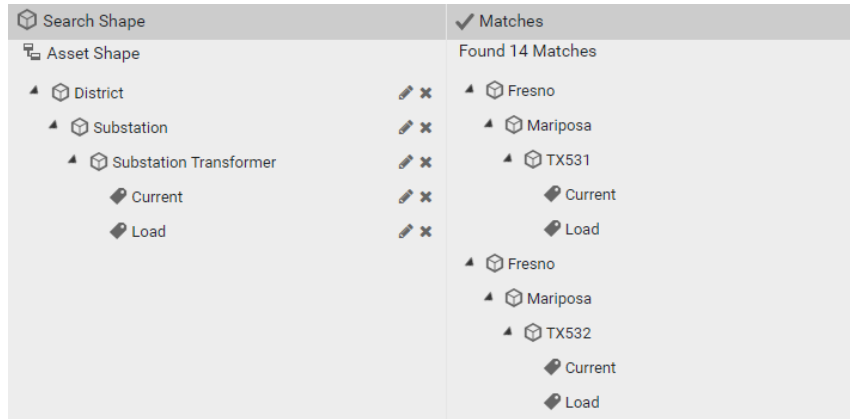
- 1) From the pancake menu, or the My Views page, create an asset view. Name it **Turbine Efficiency <Student Name>**.
- 2) From the select data page, select Create a New Shape

**Optional:** Import the shape from a preconfigured view, **Turbine Efficiency**, and skip section **3.5.3 Create an asset shape**

- 3) Select the **OSIDemo Turbine Efficiency FULL** database.

### 3.5.3 Create an asset shape


An asset shape is a pattern for searching data within the PI System. The PI Integrator will search the AF database and return each portion of the AF hierarchy that fits the asset shape as a match.



**Figure 2:** Example of an asset shape search

- 1) Pick asset Unit 1 > HRSGA in the AF Hierarchy.
- 2) Drag and drop the Unit1 element, then the HRSGA element underneath it. From the attributes section, select the attributes we defined in the data understanding section.
- 3) Drag and drop Steam Turbine1 to the asset shape, and the attribute HP Turbine Efficiency. Note this results in a single Match.
- 4) To get all the HRSGs, adjust the filter criteria in the Search Shape. Use the Edit button (pencil) next to HRSGA. Uncheck Asset Name and check Asset Template instead. This will generalize the Asset Shape to search for elements belonging to the AF template HRSG instead of the asset name HRSGA.

**Note:** Now two matches are returned.

 <b>Best Practice</b>	<p>AF templates are the ideal search filter, since all elements will resemble each other. If you need to create a shape with more disparate elements, using the search category is a good way to go. Create a new category for the asset shape if need be.</p> <p>Avoid searching by asset name with wildcards since it is expensive.</p>
---	---

- 5) To get both Combined Cycle Units, repeat the same procedure as above but for the Unit1 element, selecting the Combined Cycle Unit template.

**Note:** Now 4 matches are returned.




6) Click next to proceed to the Modify View page.


### 3.5.4 Modify the asset view's publication parameters

- 1) We can now modify our columns to match the criteria we defined in the data understanding step.
- 2) Change the start time to 1-jan-2018 00:00:00 and the end time to 1-jan-2019 00:00:00 to obtain the data for the year of 2018.

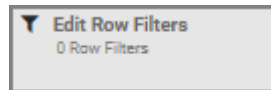
**Discussion:** What is a potential issue of using a start time of 00:00:00 with this dataset?

 <p>Tip</p>	<p>The start time determines the timestamps of the rest of the view. For example, using a start time of * (current time) will likely yield milliseconds, which are undesirable in most cases. Instead, use an absolute time such as 1-jan-2018 12:00:00 AM.</p> <p>When using a key column, data will be sampled at the timestamps of one selected attribute.</p>
--	---

- 3) Change the temperature columns to bring in the average, and rename them to spell out Temperature Average in full instead of temp.

 Tip	<p>Summary calculations are calculated over the period of the sampling frequency.</p> <p>Total calculations behave differently. AF attributes should generally be configured as TimeRange=NotSupported, or they will use the relative time configured in AF. Instead, bring in the value of the tag and use the summary functionality of the PI Integrator for BA</p>
--	---

- 4) Use the Edit Value Mode button to select a 5minute interpolation interval.




- 5) Click the Edit Row Filters button
- 6) Add a numeric row filter to remove the pressure values over 2000 psia for SH Steam Outlet Pressure since those are instrument errors.
- 7) Click the next button to proceed to the Publish screen.
- 8) Select our previously configured target or choose the pre-configured SQL target.
- 9) Publish the view.

### 3.5.5 Monitoring the view

Where does a published data reside? Since we selected the SQL target, the PI Integrator creates a table in the destination with the same name given to the view on creation.

- 1) To see the table in SQL, open Microsoft SQL Server Manger Studio 2014 (SSMS) and connect to PISRV01 with Windows Authentication:
- 2) Find the table with the view name you chose (**Turbine Efficiency <Student Name>**) under Databases-> SQLDatabase-> Tables.

 Tip	<p>If you wish to maintain the AF security model on your data, publish the data to a PI View. Then use the PI ODBC client on the client machine to connect to PI SQL DAS that distributes PI View data according to the view permissions set in the PI Integrator UI.</p>
--	---

There are also tools in the PI Integrator UI to ensure that publication is running smoothly.

- 3) Click on the logs tab to view the logs related to this view. You can enter a start time and end time, filter for different levels of log messages, and export the logs to a .csv file.
- 4) Click on the statistics tab to see publication statistics. The number of rows written as well as the error count is displayed in tabular format as well as in a historical trend for ad-hoc analysis.

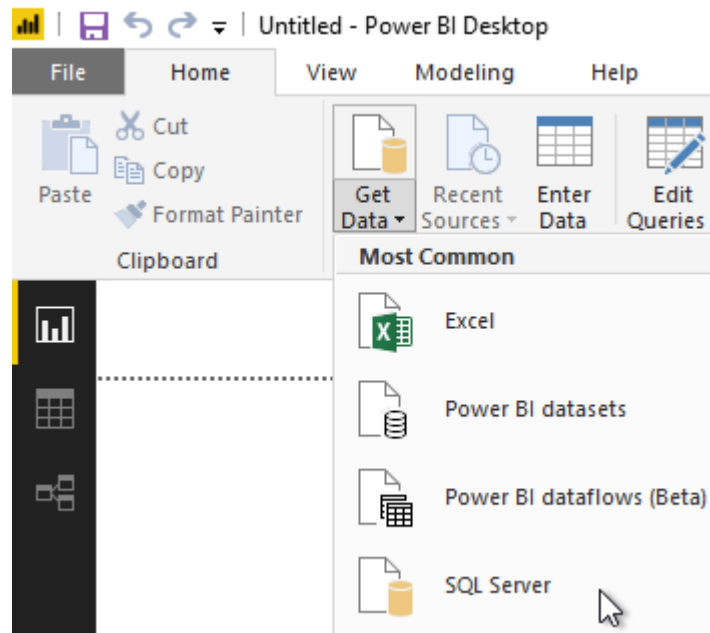




Tip

The logs of the view only show the log messages related to that view. The logs in the administration section show the log messages for the entire PI Integrator instance.

### 3.5.6 Viewing the data in a BI tool

- 1) Open Power BI desktop using the shortcut on the desktop. From the Home ribbon, expand the Get Data button and pick SQL Server.



- 2) Connect to the server PISRV01
- 3) Expand SQLDatabase and load your table **Turbine Efficiency** <Student name>
- 4) Add a Slicer  to the report canvas and drag the Combined Cycle Unit column from the data set into the Field box to create a slicer for the units
- 5) Add a Scatter Plot  and drag the following columns to the corresponding fields:
  - **X Axis:** Cold Reheater Steam Temperature Average
  - **Y Axis:** HP Turbine Efficiency
  - **Legend:** Combined Cycle Unit



## 4. Directed Activity – Well Drilling - Stick-slip events

### 4.1 Objective of Activity

Examine a well drilling site and follow the first 3 steps of the CRISP-DM methodology to generate a dataset for further analysis.

### 4.2 Tasks

- Perform the business understanding step of the CRISP-DM methodology
- Perform the data understanding step of the CRISP-DM methodology
- Perform the data preparation step of the CRISP-DM and generate a dataset using an event view

### 4.3 Business understanding

#### 4.3.1 Scenario

Stick-slip conditions happen due to the friction at the interface between two materials. In the well drilling industry, the most dangerous stick-slip condition occurs at the interface between the end of the drill string and the rock. The interface stops moving due to the friction, but the torque is continually applied. The system stores energy and eventually releases it at once, causing the drill bit to turn several turns. This sudden slip out of control is associated with destructive vibrations and is responsible for a significant amount of failures in the industry.

Stick-slip can induce harmonic oscillations in the drill string or can happen between the hole wall along the length of the drill-string. In this example, we will be exporting data from the stick-slip events from different wells in a site so that they can be submitted for further analysis by experts.



Figure: Well drill string (<https://www.tomax.no/resources/about-stick-slip/>)

### 4.3.2 Produce a project plan

**Discussion:** Who are the stakeholders in such a project? What kind of data can we look at?

The well site manager, the process engineer, the well operators, well vendors.

- **Inputs:** Use the data in the PI System.
- **Analysis:** Well rig vendors have their own analysis tools that can analyze the well rig data.
- **Output A:** Produce a table of all the stick-slip events to perform preliminary analysis can be performed in Power BI.
- **Output B:** Send granular data to the well rig vendor for analysis by their specialists.

### 4.3.3 Identify business success criteria

**Discussion:** Based on our discussion so far, what could be the business success criteria?

**Solution:** Reduce the frequency and severity of stick-slip events.

## 4.4 Data understanding

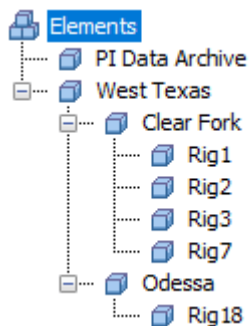
### 4.4.1 Data Exploration

Let's explore the data we have available to us in the AF Database. Find all the attributes that will be helpful for analysis.

Log on to the server PISRV01 and open PI System Explorer. Connect to the following AF Server and AF database:

**AF Server:** PISRV01

**AF Database:** OSIDemo UOG Well Drilling & Completion Monitoring FULL



There are two sites in West Texas, Clear Fork and Odessa. Rigs 1, 2, 3 and 7 are installed in Clear Fork and rig 18 is in Odessa.

Take 5 minutes to explore the AF Hierarchy and find relevant attributes based on the requirements in the business understanding section above.

The process engineer told us that we can define a stick-slip event by the following conditions:

- The well is in drilling state.
- The depth is more than 1 ft
- The change in depth in the past 2 hours is less than 0.1 ft
- The top drive RPM is higher than 15.

**Discussion:** what is the difference between the data from this scenario and the last scenario?

**Solution:** We do not need all the data recorded, only from the events under specific conditions. In the PI system, we can store these events as event frames.

- These event frames have been created in the AF database under the event frame template stick-slip. The event frames have average calculations attributes configured.
- Stick-slip events last around one hour
- They are likely to happen at a depth of 6,000+ ft. We can safely ignore the ones happening at a depth of less than 150 ft, because those are known startup problems.

These tags have the following properties:

- Process tags update every 3 seconds.
- The well rig vendor can process all the events at this frequency.
- Our data scientist would like to start with the data for the year of 2018, as a first pass analysis.

## 4.5 Data Preparation

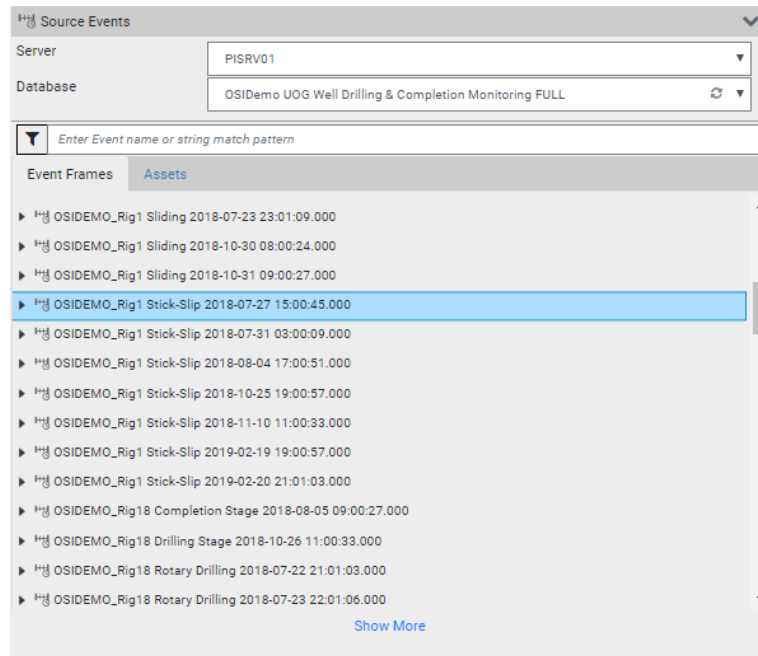
We will be publishing the data to the same target destination as in the previous exercise, so there is no need to create a new target in the PI Integrator for Business Analytics.

### 4.5.1 Create a summarized event view

We can create an event view to export the summarized stick-slip event frames (output requirement A).

- 1) From the pancake menu, or the My Views page, create an event view. Name it **Stick Slip <Student Name>**.
- 2) Create an event shape, or import the shape from the preconfigured view, Stick Slip.
- 3) Select the **OSIDemo UOG Well Drilling & Completion Monitoring FULL** database.


- A list of recently generated event frames appears in the source events section. Scroll down to find a Stick-Slip event. It follows the naming convention **OSIDEMO\_Rig# Stick-Slip <Starttime>**



**Optional:** Use the event filter  option to search for event frames of template Stick-Slip.

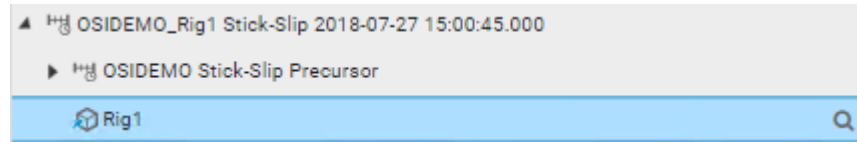
**Note:** You may notice that there are some precursor event frames that use the same template. We will filter them out in the following steps.

- Drag and drop a stick-slip event frame to the event shape, then edit the filter criteria to search for events from the stick-slip template. (Click the pencil icon to edit the filters)

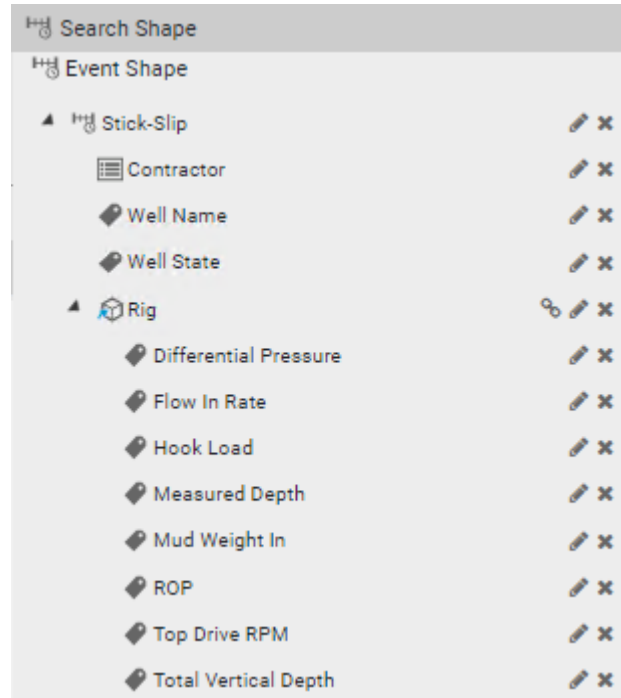
 <b>Best Practice</b>	<p>AF templates are the ideal search filter, since all events will share similar attributes. If you need to create a shape with more disparate elements, try searching by AF categories. Create a new category for the event shape if need be.</p> <p>Avoid searching by event name with wildcards since it is expensive.</p>
---	---

- With the event frame selected, select the Contractor, Well Name and Well State attributes and click and drag them to the event shape. These are attributes from the event frame.



- 7) We are interested in the process variables defined the rigs' element attributes as well, because only a few summary calculations were created in the event frame attributes. Expand the event frame to find its child event frames as well as its referenced element.



- 8) Select the rig and build the following shape.



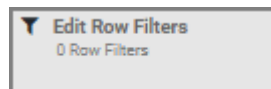
**Note:** Remember to edit the element search filter to search for the rig element template!

 Tip	If additional element attributes are required from elements that are related to the referenced asset, use the  button to unlink the asset shape from the event shape, which will allow you to build a complex asset shape.
--	---

- 9) Click next to go the modify view page.


#### 4.5.2 Modify the Event View's publication parameters

- 1) Change the start time to 1-jan-2018 00:00:00 and the end time to 1-jan-2019 00:00:00, as this will be a first pass analysis.



- 2) Click the Edit Row Filters button



- 3) Select the string row filter and configure it so that the name of the event frame is not equal to OSIDEMO Stick-Slip Precursor. This will filter out the precursor child event frames automatically generated by PI Analysis.
- 4) Add a row filter to remove the event frames that occur at a total depth of less than 150ft, since those are startup issues that we are not concerned about for this project.  
**Hint:** Add a numeric filter on the total measured depth row
- 5) Modify the Flow In Rate, Hook Load and Mud weight In columns to be averages, and rename the columns appropriately.
- 6) Publish the view in run once mode, using the SQL target.

 <b>Tip</b>	<p>The PI Integrator for BA performs summary calculations are done over the entire duration of the event frame even in sampled value mode.</p> <p>PI AF event frame attributes can be used to retrieve values in different methods, or PI AF calculations can be used to create calculations in different modes.</p> <p>Consider performing calculations downstream once the dataset is published.</p>
---	--

#### 4.5.3 Monitoring the view

The same procedure as above can be followed to check the publication for errors and for statistics. Connecting to the SQL database will allow us to view the data table.

#### 4.5.4 Viewing the data in a BI tool

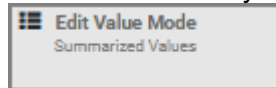
- 1) Connect to the table using Power BI and connect to the SQL database as a data source again. Import your table called **Stick-Slip <Student Name>**
- 2) Add a Slicer  to the report canvas and drag the Well Name column from the data set into the Field box to create a slicer for the wells.
- 3) Add a Scatter Plot  and drag the following columns to the corresponding fields:
  - **X Axis:** Top Drive RPM
  - **Y Axis:** Total Vertical Depth
  - **Legend:** Well Name

**Discussion:** What kind of preliminary conclusion can we find from a quick preliminary look at the data?

#### 4.5.5 Create a dataset with granular data

We will create a copy of the event view to export the granular data contained in the slip event frames for further analysis (output requirement B).

- 1) From the My Views page, select the event view **Stick-Slip <Student Name>** that has just been published. Use the modify view button to create a copy of the view. Name it **Stick Slip <Student Name> 3s**.
- 2) Go to the Modify View page. Use the Edit Value Mode button



to change the value mode to sampled values, interpolating every 3 seconds, which corresponds to the resolution of our tags.

**Discussion:** Will it make a difference if we used a key column instead?

- 3) Publish the view to SQL.

**Discussion:** In a real-life scenario, vibration data would likely come at a much higher frequency. What kind of consideration should we take when choosing how we export this data?

## 4.6 Modelling, Evaluation and deployment

### 4.6.1 Modelling

Now that we have a better idea of what a dataset looks like, what can be done with it? Modeling can be summarized in 3 steps:

- 1) Sample data
- 2) Learn a model
- 3) Make predictions

#### Sample data

This step has been completed in the data preparation step.

#### Learn a model

Part of the dataset is then used as the training data to find the relationship between the parameters, which can then be scored against a different subset of the data.

## Make predictions

A simple model could be a mathematical equation that can be recreated in PI Analysis, but a more complex model can be thought of a black box which accepts specific inputs (process variables) and generates an output (the predicted value of the parameter of interest). New data can then be sent to the model, and predictions are made.



### 4.6.2 Evaluation, deployment

This step assesses the degree to which the model meets the business objectives and seeks to determine if there is some business reason why this model is deficient. Does the model generated help us solve our business objective? If so, proceed to testing or deployment. If not, overcome the challenges identified and repeat the process.

The predicted values can be operationalized by sending them back into the PI System and building dashboards, alerts, or patterns can be used to make adjustments to operations.

In the Heat Recovery Steam Generators example, a potential outcome could have been to set up alerts for monitoring the feed water flow rate and temperature, or to install an additional filter to remove impurities from the feedwater to prevent boiler performance degradation.

In the Well Rig operation example, a potential outcome would be to reduce the rotational speed at depths of 6k-8k ft in the Clear Fork site, or to increase the mud weight during operation.

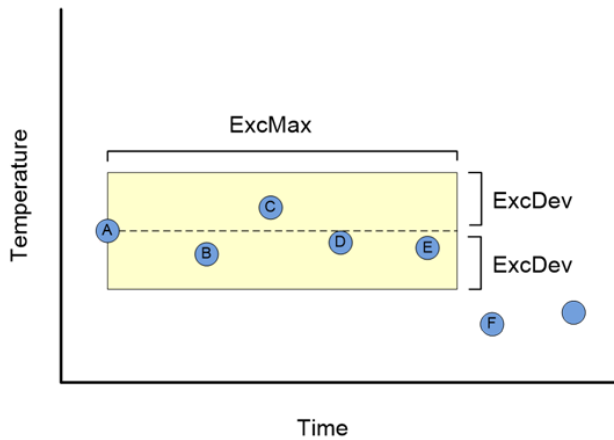


# Appendix A

## Sampling interval – Basics

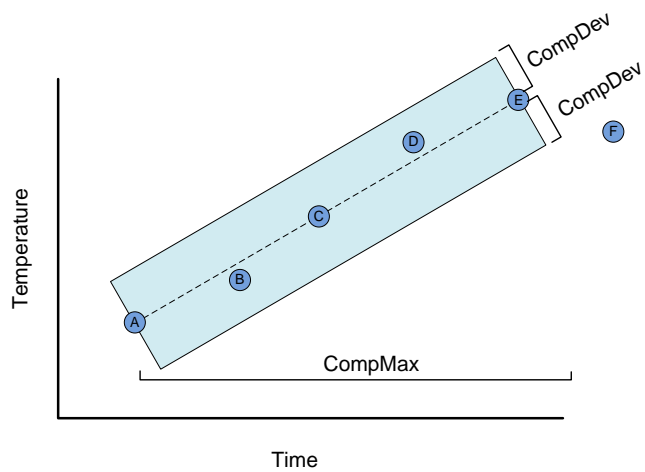
### Exception reporting

Exception report is used to filter out noise in the instrumentation. If the value does not change by a significant amount, do not send the value to the PI System.



### Compression testing:

Store only significant values. If the slope is constant, the value does not need to be saved.

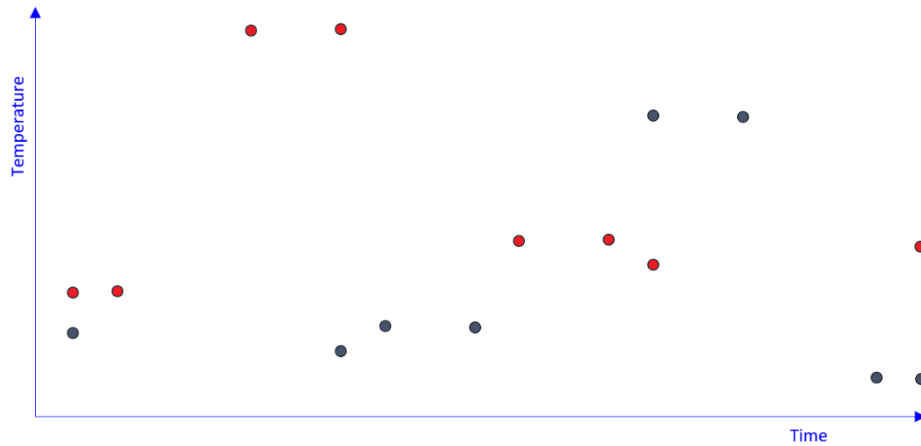


## Scan class and offset

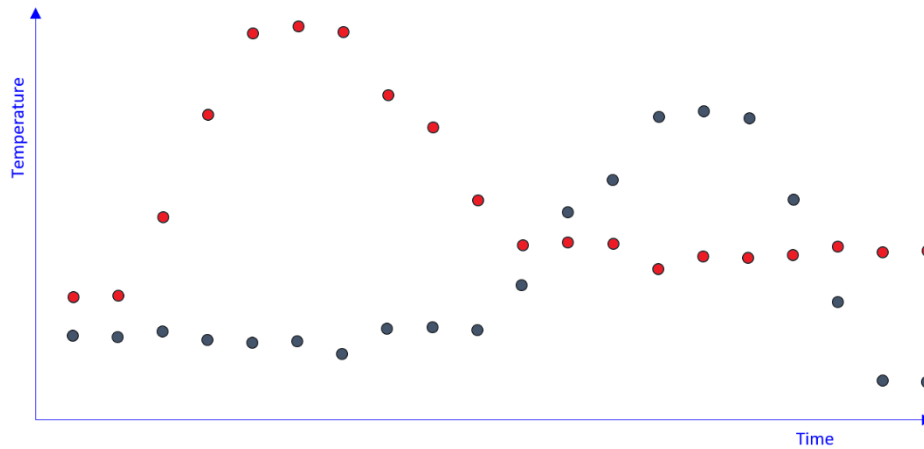
The scan class determines the frequency at which input tags are scanned for new values. If an offset is specified, scan time is calculated from midnight on the day that the interface was started, applying any offset specified.

For example,  $/f=00:01:00,00:00:05$  means that we scan for a new value at the 5<sup>th</sup> second of every minute.

Depending on Exception, Compression, Scan class and offset settings, it is still possible that values were scanned at the same time for two tags that do not have data saved at the same timestamps.



Data after exception and compression



Data before exception and compression

## **Non-uniform data**

Which column is more important? If there is a column that is more important, use key-column sampling. This will sample all other columns at the timestamps of the key column.

For example, if an expensive, time consuming or manual measurement is only made occasionally, using that column as the key column will reduce the amount of data exported.

For higher frequency data, consider using averages to minimize the effects of down-sampling.

## Appendix B

### Five parameters are relevant when configuring a Continuous View.

On the initial publication,

- Start Time (ST): Start time of the initial publication of the view
- End Time (ET): End time of the initial publication of the view
- Sample Frequency (SF): Sampling time interval between each row published
- First run (FR): When the view will first run

On each scheduled publication thereafter,

- Start Time (ST): Last published End Time + Sample Frequency
- End Time (ET): Configured end time, evaluated at run-time, for example \*-15m is 15 minutes ago.
- Sample Frequency (SF): Sampling time interval between each row published
- Run Frequency (RF): Period of time between each scheduled publication

For more detail, consult [OSIsoft KB01683 - Understanding PI Integrator View Scheduling](#)



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

