

Time-series data is different, but not hard

Gopal GopalKrishnan, P.E.

16th April, 2015



ABSTRACT



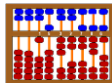

Big data has been dominating the IT mindshare for the last few years. And, with a slew of new technologies, i.e. in-memory database, NoSQL, machine learning, and cloud added to the Big data - Hadoop ecosystem, IT is left to grapple with reality vs. hype. Where do you begin, and how do you proceed? McKinsey, in a January 2015 article – [“Getting Big Impact from Big Data”](#) - makes these recommendations:

“Visualization tools... are putting business users in control of the analytics tools by making it easy to slice and dice data, define the data exploration needed to address the business issues, and support decision making,” writes David Court. Earlier in the article he says “...analytics specialists builds models targeted to specific use cases. These models have a clear business focus and can be implemented swiftly.”

Sensor data, i.e. time-series data from aircraft or ground assets or from manufacturing operations, has its own flavor of bigness – along with its three V’s - volume, velocity, and variety to make up the industrial big data. But we have successfully dealt with it across numerous industries – in power generation, oil & gas, chemicals, pharmaceuticals, metals and mining, paper & pulp, utilities such as water, gas and electric transmission and distribution, critical facilities, data centers, and others.

And, in working with our customers, we find self-service data analytics using models targeted to specific use cases is key to rapid insights – whether it is small data or large data, or even big data. Attend this session – we will explore the various ways to **getting insights from data**.

Time-series data is different – but not hard

- Instruments and sensors - data, data sampling, data rates, data compression... 
- Shaping the data - list, hierarchical..., data structure, data dictionary... 
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT... streaming data 
- Framing the time-series data - Start/End of a time window... 
- Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

Time-series data is different – but not hard

- **Instruments and sensors** - data, data sampling, data rates, data compression...
- Shaping the data - list, hierarchical..., data structure, data dictionary...
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT... streaming data
- Framing the time-series data - Start/End of a time window...
- Out-of-the-box - visualization, display tricks, Excel, Notifications, smart device...
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

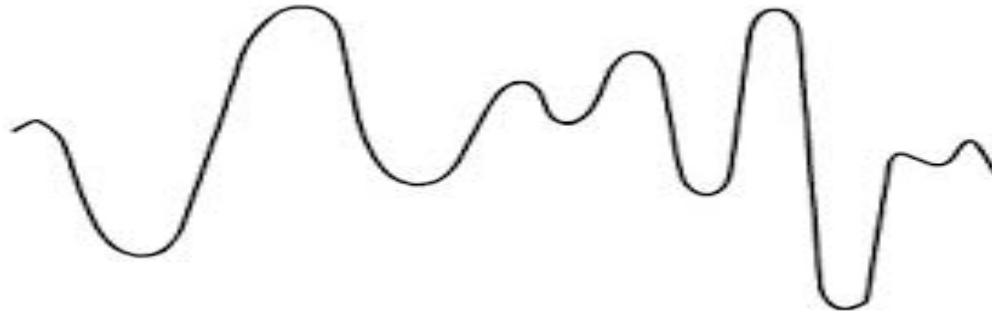
Time-series data is different

File store API available incomplete data Rows not aligned
MySensor 3.14 04-16-2015 01:10:27 PM

1429189827 sec POSIX UTC

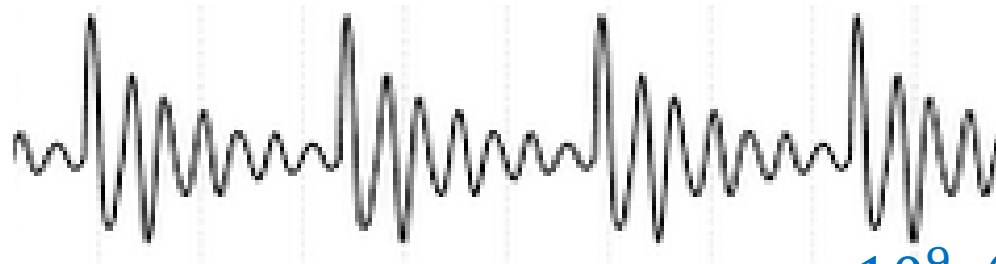
1429189827.35897 sec

~ 15 microseconds precision



Dynamic Sampling $10 \mu\text{Hz} \sim 50 \text{ kHz}$

once a day to microseconds



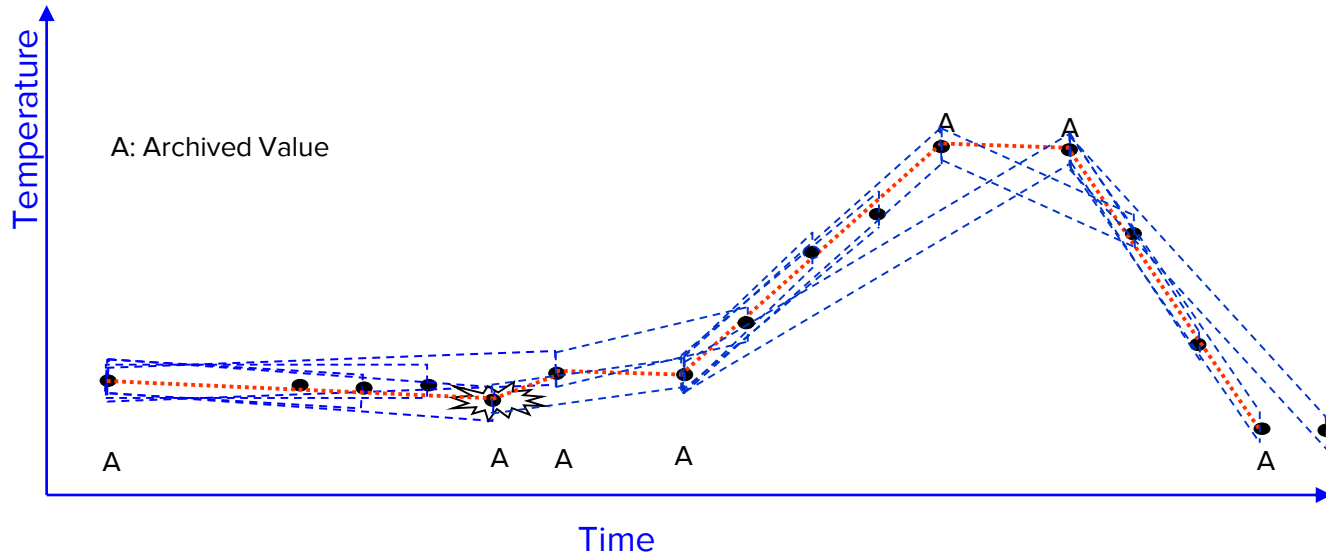
10^{-5} to 10^4

signal frequency range

10^9 (billion) order of magnitude

Data compression

~ 15 fold



Time-series data is different – but not hard

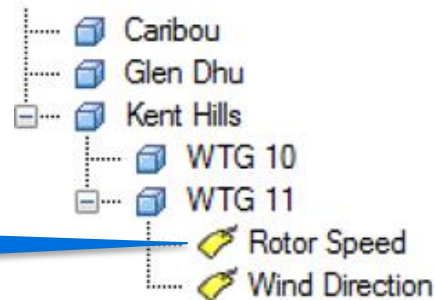
- Instruments and sensors – data, data sampling, data rates, data compression...
- **Shaping the data** - **list, hierarchical..., data structure, data dictionary...**
- Calculations – virtual sensor, expressions, roll-ups, SQC, FFT... streaming data
- Framing the time-series data - Start/End of a time window...
- Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

Specifying the meta-data



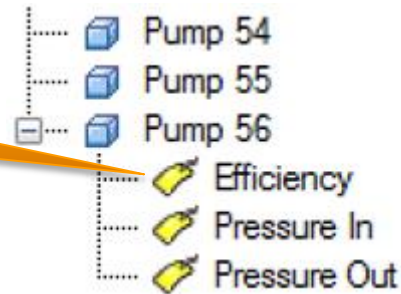
PI AF (Microsoft SQL)

GT56.TIC.PV

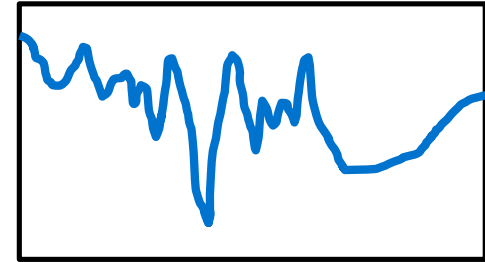


WT11.SI.PV

P56.PEF.CALC



GP23.ATHK8.PV



	A	B
1	Efficiency	74.54%
2		
3		



Rotor Speed



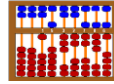
Efficiency

Time-series data is different – but not hard

- Sensor - data sampling, data rates, data compression...
- Shaping the data (list, hierarchical...) - data structure, data dictionary...
- **Calculations – virtual sensor, expressions, roll-ups, SQC, FFT... streaming data**
- Framing the time-series data - Start/End of a time window...
- Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

Calculations – virtual sensor, expressions, roll-ups, SQC, FFT... streaming data

Calculation – Expression



Boiler Efficiency = Average(Boiler1, Boiler2...BoilerN)

Boiler1

Flow Out

Fuel Flow Rate

Efficiency = (Flow Out / Fuel Flow Rate * 3.14)



Boiler Template

Boiler2

Flow Out

Fuel Flow Rate

Efficiency

Boiler3

Flow Out

Fuel Flow Rate

Efficiency

Calculation – Rollup



Elements

- Generation
 - OSISoft Power
 - Big Creek Power Plant
 - Cleveland Power Plant
 - Houston Power Plant
 - Philadelphia Power Plant
 - San Leandro Power Plant
 - Unit 1
 - Air Heater
 - Balance of Plant
 - Feedwater System
 - Boiler Feed Pump #1
 - Boiler Feed Pump #2
 - Generator
 - Mills
 - Turbine
 - zz_RunHours_Report
- Wind Power Generation Fleet
 - Transmission and Distribution

- Element Searches
- zz_RunHours_Report
- Wind Power Generation Fleet
- Transmission and Distribution
- Element Searches

Elements

- Event Frames
- Library
- Unit of Measure
- MyPI
- Notifications
- Contacts
- Analyses

Boiler Feed Pump #1

General Child Elements Attributes Ports Analyses Version

Boiler Feed Pump #1

General Child Elements Attributes Ports Analyses Version

Name	Backfilling
Boiler Feed Pump High Bearing Temperature	<input checked="" type="checkbox"/>
Boiler Feed Pump Low Discharge Flow Anomaly	<input checked="" type="checkbox"/>
Boiler Feed Pump Low Pump Speed	<input checked="" type="checkbox"/>
Boiler Feed Pump Suction Pressure Anomaly	<input checked="" type="checkbox"/>
Boiler Feed Pump Vibration Anomaly	<input checked="" type="checkbox"/>

Name: Boiler

Description:

Categories:

Analysis Type: ☐ External

Event Frame Template:

Name	Expression	Value
Outboard Y' Bearing Vibration Y' > Limit	if InboardXFault or InboardYFault or OutboardXFault	False

Evaluated at 3/29/2015 7:10:46 PM

StartTrigger true for: Minutes

☒ Generate child root cause event frame before parent event frame starts

Duration: Days

Name: Root Cause

Category:

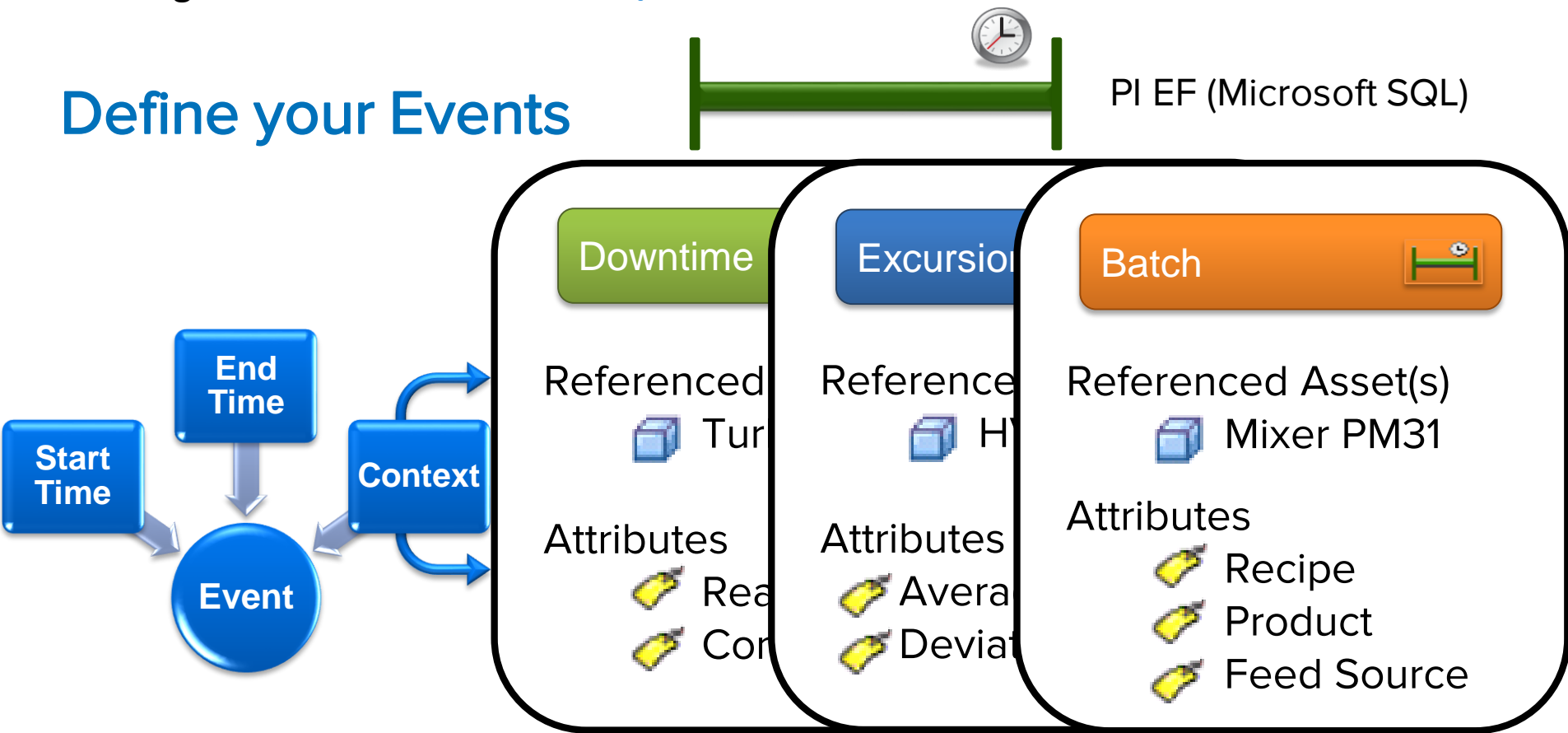
Scheduling: ☒ Event-Triggered ☐ Periodic

Trigger on: Any Input

Time-series data is different – but not hard

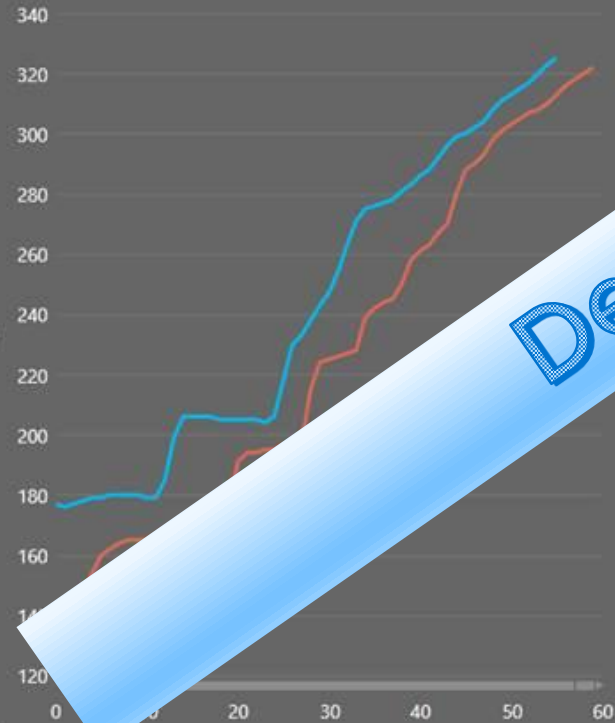
- Sensor - data sampling, data rates, data compression...
- Shaping the data (list, hierarchical...) - data structure, data dictionary...
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT... streaming data
- **Framing the time-series data - Start/End of a time window...**
- Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

Define your Events



Digester - Heating Phase

Heat-Temp-Top by TimeAfter, and EventFrame



Demo

EF

DigCyc

20150101

20150102

20150103

20150104

20150105

20150106

20150107

20150108

20150109

20150110

20150111

20150112

20150113

20150114

20150115

20150116

20150117

20150118

20150119

20150120

20150121

20150122

20150123

20150124

20150125

20150126

20150127

20150128

20150129

20150130

20150131

20150201

20150202

20150203

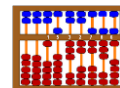


OSIsoft.

FEDERAL WORKSHOP

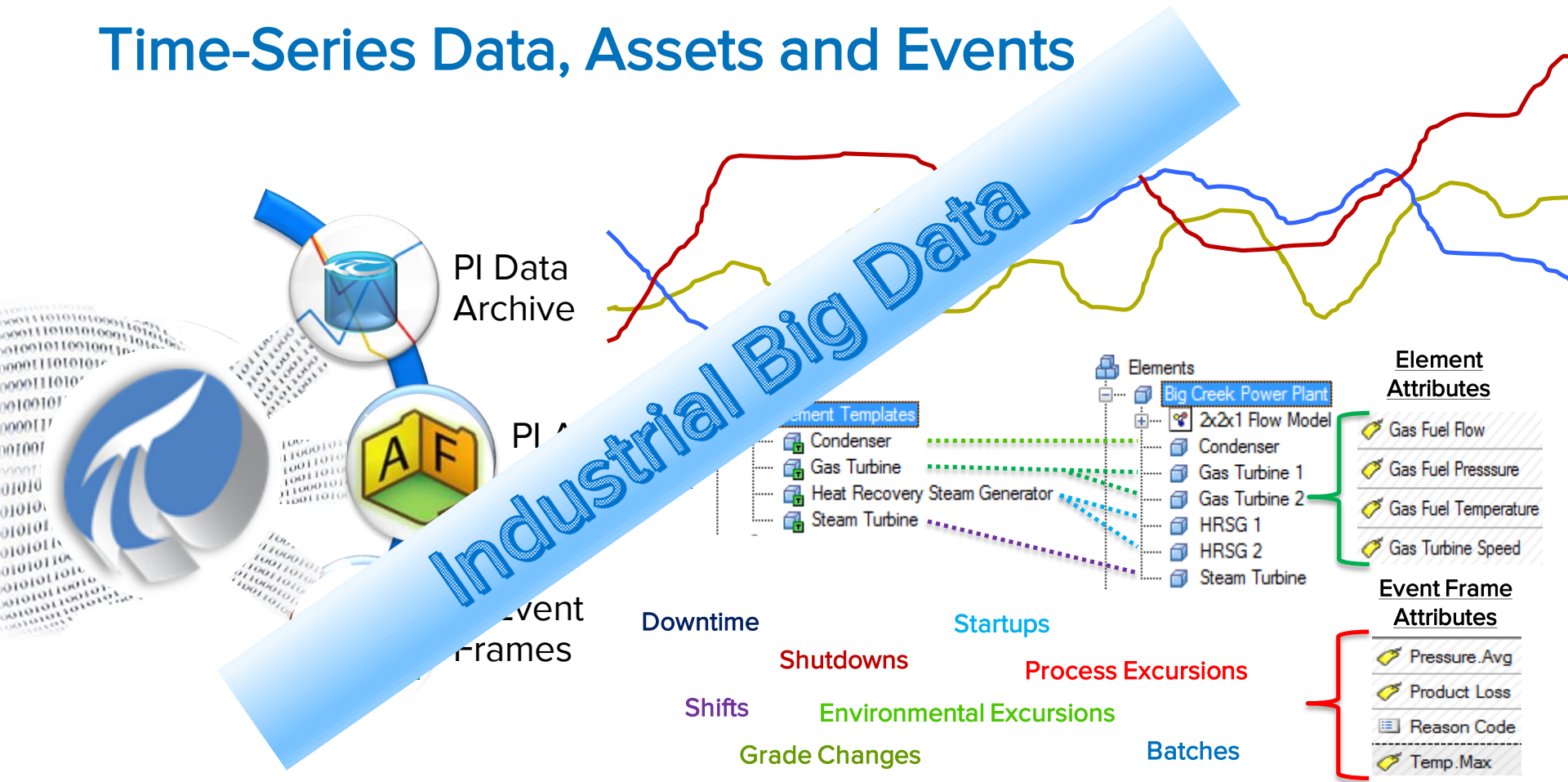
Time-series data is different – but not hard

- Instruments and sensors – data, data sampling, data rates, data compression...
- Shaping the data - list, hierarchical..., data structure, data dictionary...
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT... streaming data
- Framing the time-series data - Start/End of a time window...



- Out-of-the-box - visualization, display tricks, replay, Excel, Notifications, smart device...
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

Time-Series Data, Assets and Events



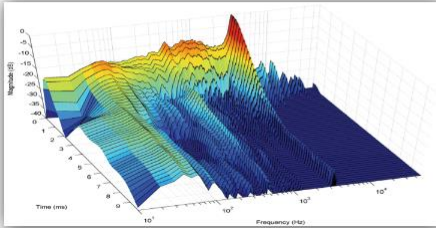
OSIsoft.

FEDERAL WORKSHOP

Industrial Big Data

- Volume - **huge**
 - Millions and billions and trillions of **measurements**
- Velocity - **high**
 - Every few seconds, and some with millisecond/microsecond precision
- Variety - **many**
 - Production - **sensor, automation controls - time-series data**
 - Quality - **lab data**
 - Maintenance - **relational, text**
 - Inventory - **relational, transactions**
 - Other - **manual input with annotations, web pages, external systems...**

PI SERVER 2015



Syncro Phasors

4.8K data streams, 120Hz
3 years online
Unique Events: 55 Trillion
Estimated Data: 430TB

430TB



Data Center

100K cells, 2M breakers
10 years online
Unique Events: 105 Trillion
Estimated Data: 840TB

840TB



Automated Metering

20M meters, 5-min reads
7 years online
Unique Events: 177 Trillion
Estimated Data: 1,410TB

1,410
TB





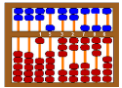

Fleet Monitoring

1K assets, 1M points
10 years online
Unique Events: 6,307 Tr
Estimated Data: 50,460TB

50,460
TB

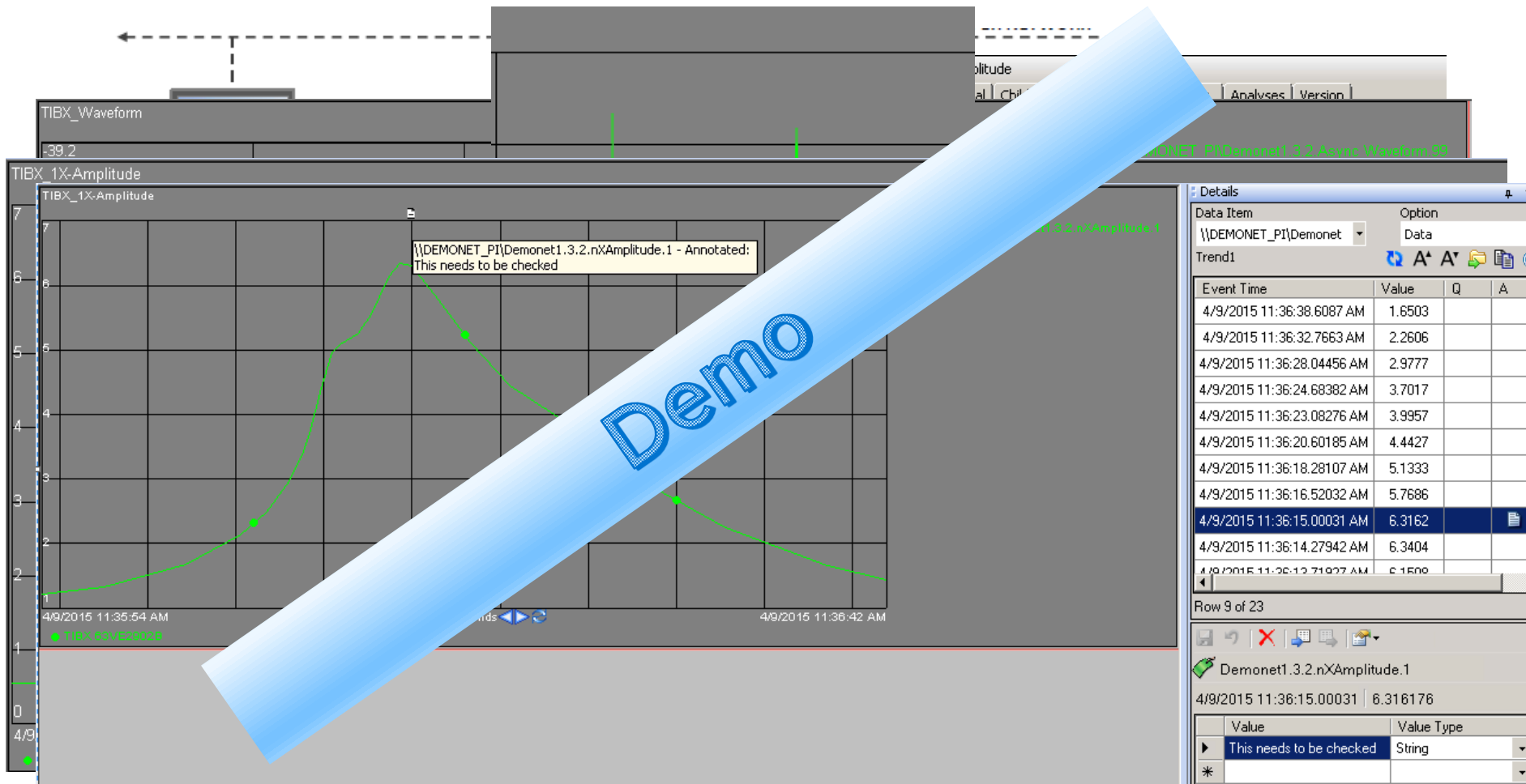


Time-series data is different – but not hard

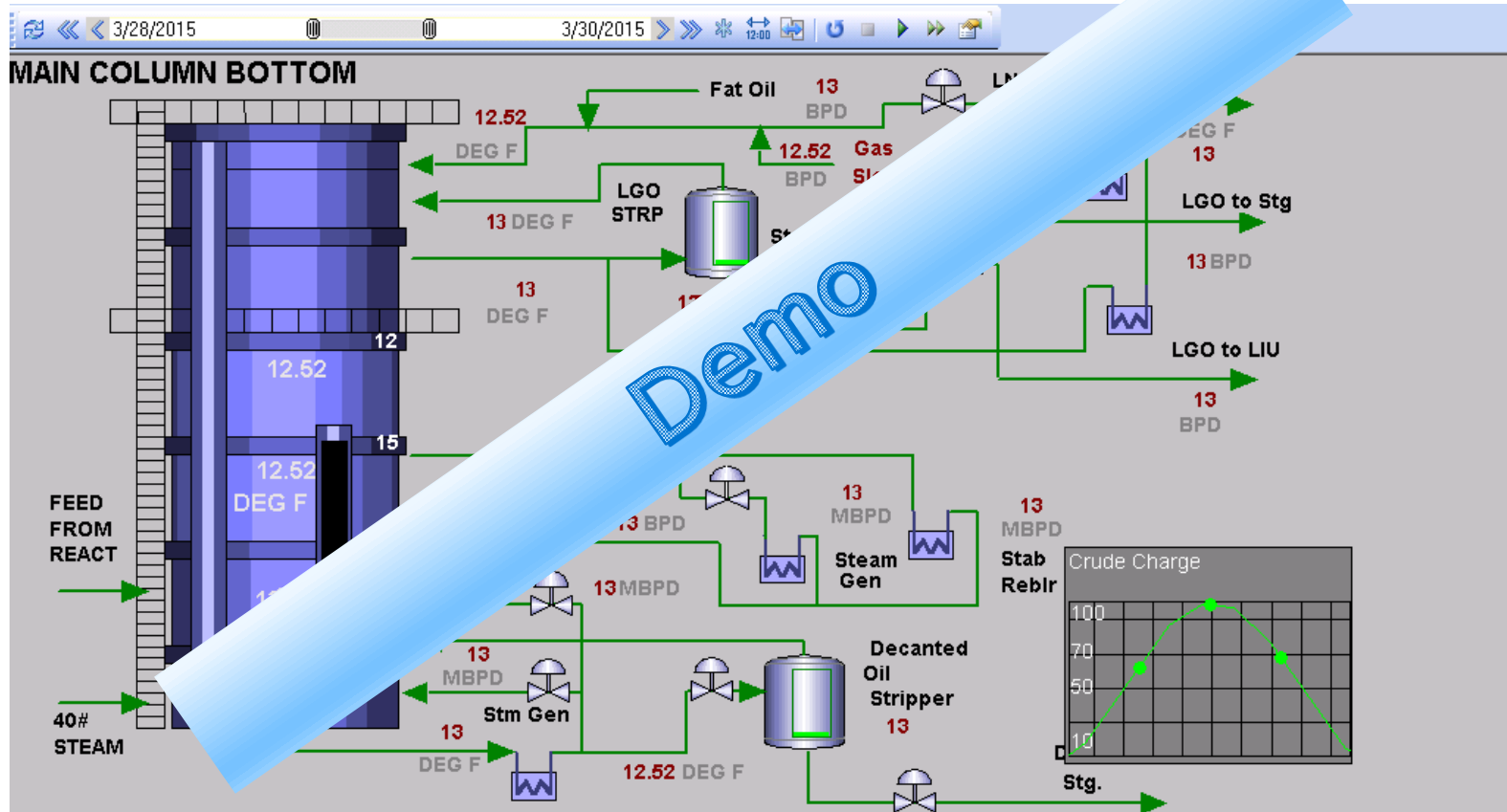
- Sensor – data, data sampling, data rates, data compression... 
- Shaping the data - list, hierarchical..., data structure, data dictionary... 
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT... streaming data 
- Framing the data - Start/End of a time window... 
- **Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...**
- External - business intelligence, data mining, machine learning, Matlab, R, Big Data...

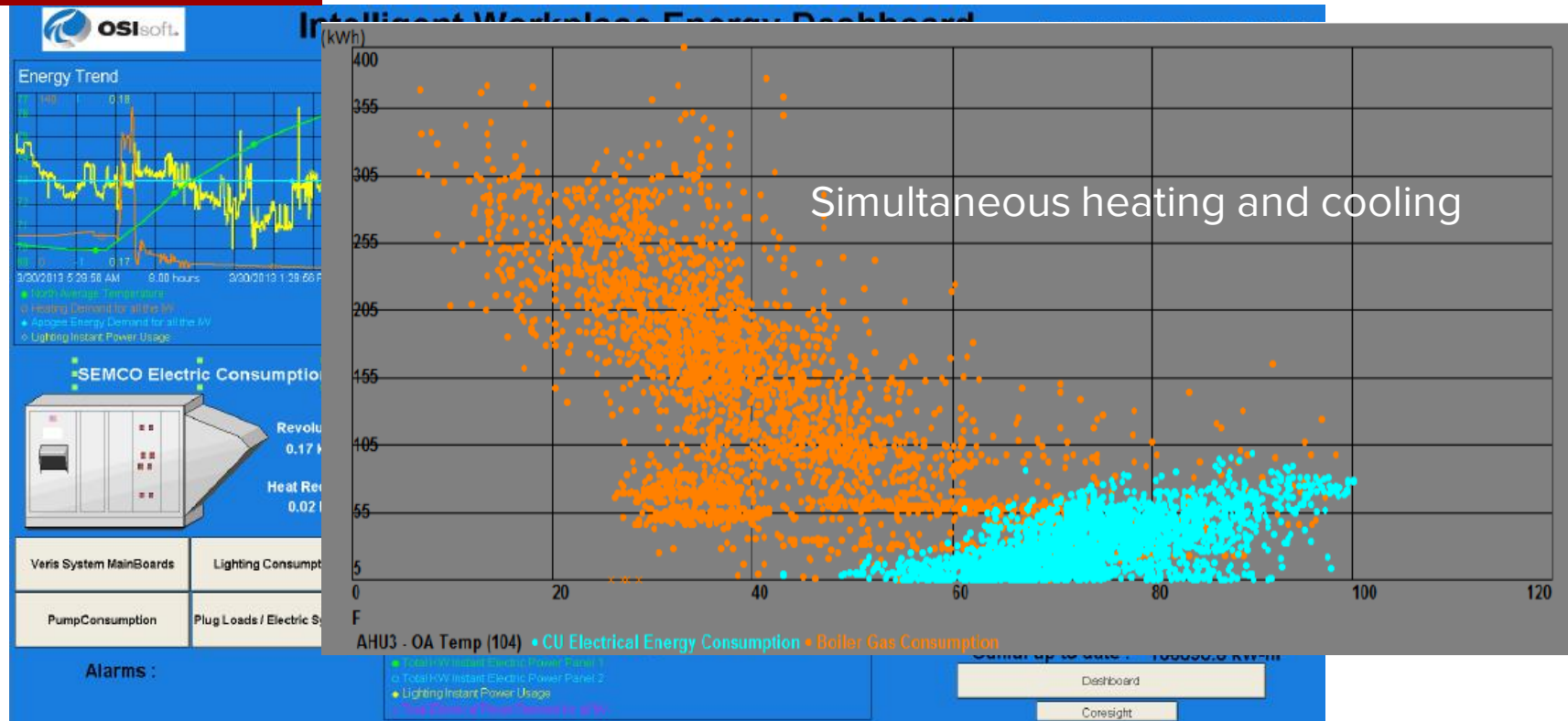
Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...





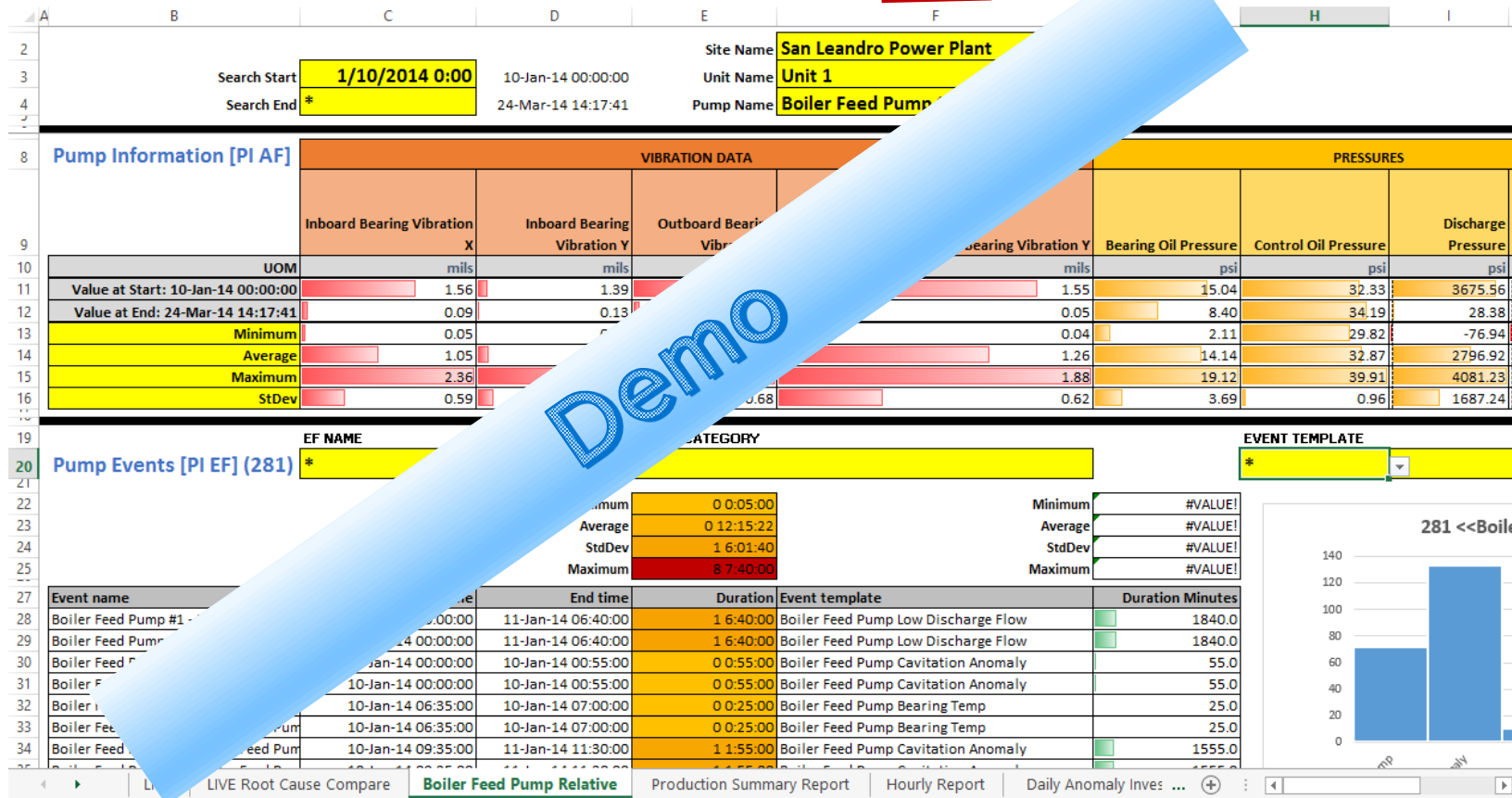
Replay time-series data





<http://www.osisoft.com/Templates/item-abstract.aspx?id=11029>

Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...



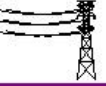
Boiler Feed Water Pumps

Analysis

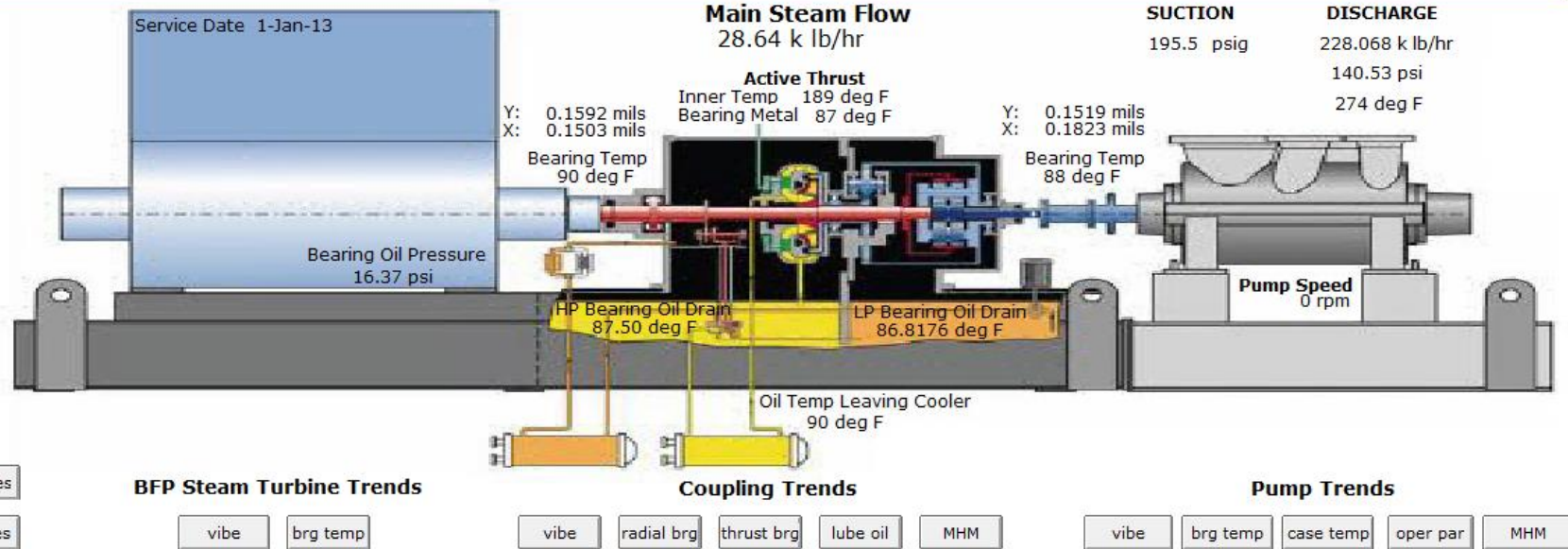
Station Ovr/Vw

Generator

2 MW gross

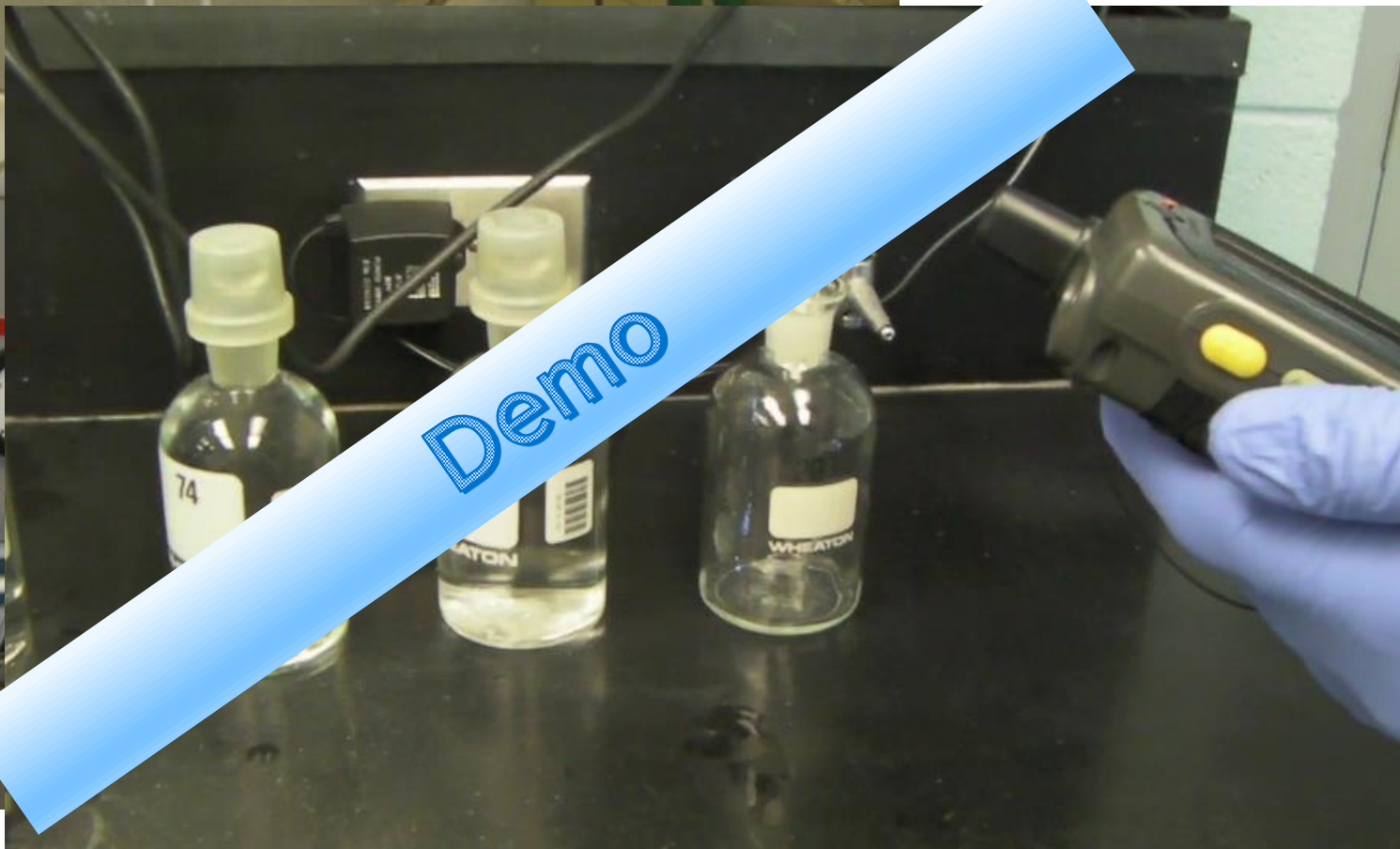


Pump #1





Demo



OSIsoft.

FEDERAL WORKSHOP

© Copyright 2015 OSIsoft, LLC

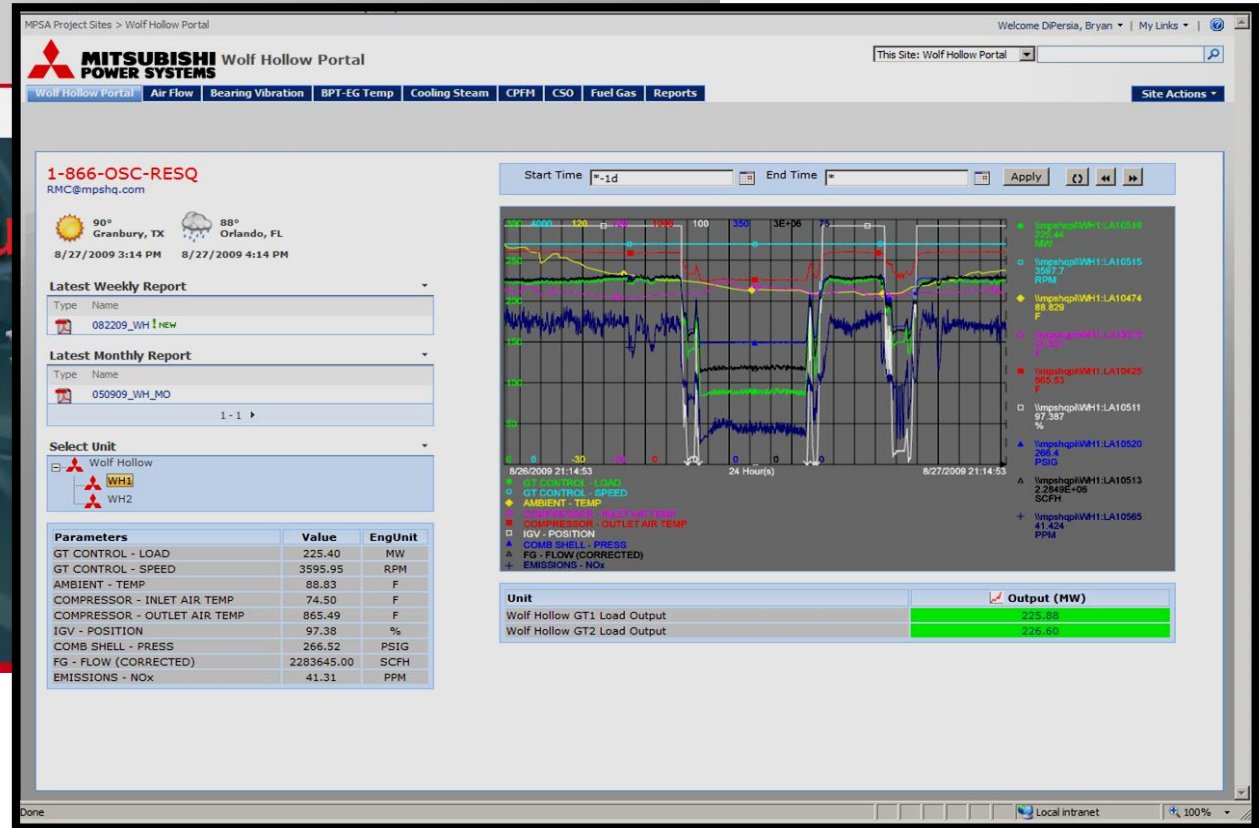
Remote monitoring – several hundred gas turbines

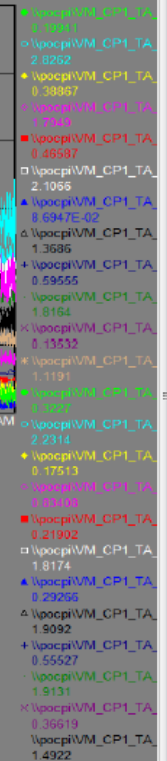


Company Products Service

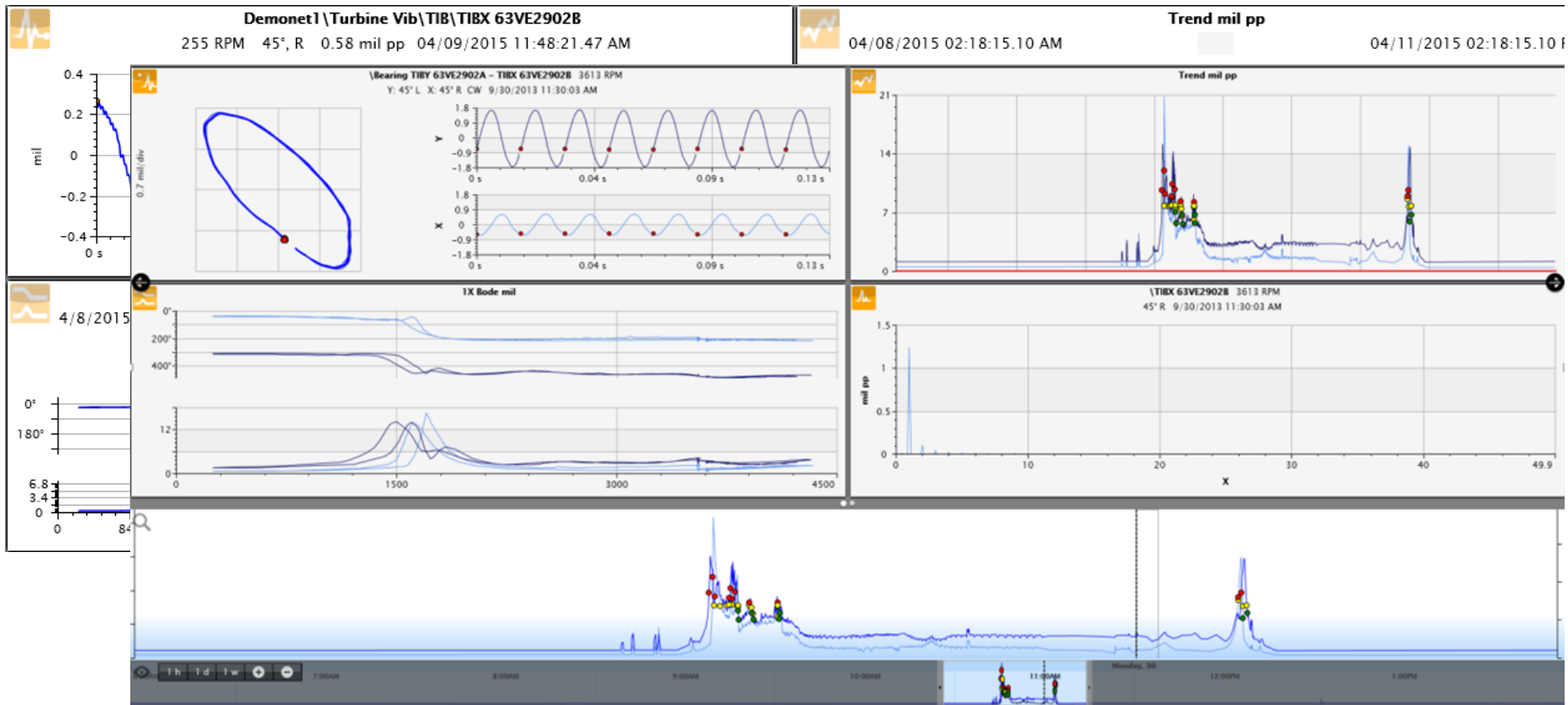
J-Series Gas Turbines
for Grand River
Dam Authority's
New Gas Fired
Generation Facility

[Learn More](#)

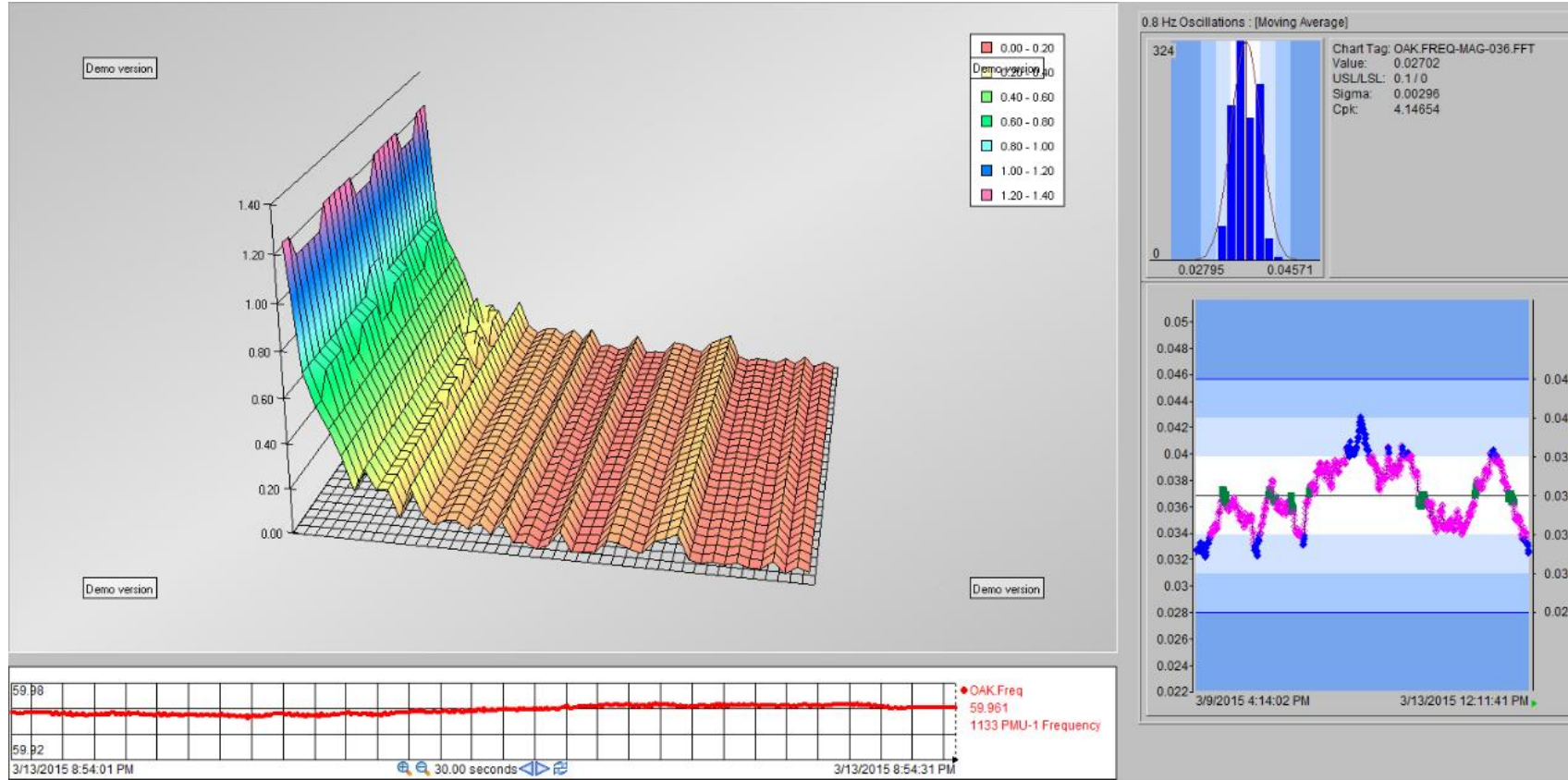




OSIsoft Partner: Metrix — SETPOINT www.setpointvibration.com



Phasor measurements - FFT and SQC charts




Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device...



Thu 3/26/2015 7:00 AM

DoNotReply@OSIsoft_PINotif.com

Flash 3 Press\Notifications[Daily FD3 Production KPIs] generated a new notification event.

To  Gopal Gopalkrishnan

Name: Daily FD3 Production KPIs

Trigger Time: 3/26/2015 6:00:00 AM Central Daylight Time (GMT-05:00:00)

Unit: Flash 3 Press

In the Last 24 hours:

F3 Press Cycles = 202 count

F3 Total Pounds Produced = {Error retrieving result} lb

F3 Downtime = 2.166667 h

Cake Wash = 56814.78 US gal

Slurry Feed = 12145.21 US gal

<http://dfpicoresight/Coresight/#/PBDisplays/10073>



OSIsoft.

FEDERAL WORKSHOP

© Copyright 2015 OSIsoft, LLC

Boiler Feed Water Pumps

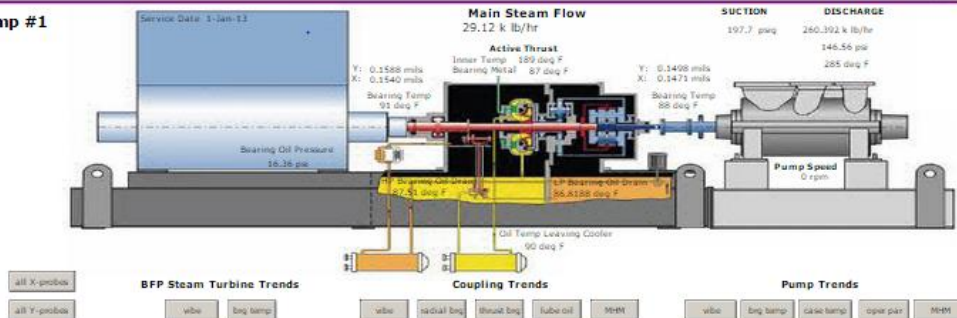
Analysis

Station Overview

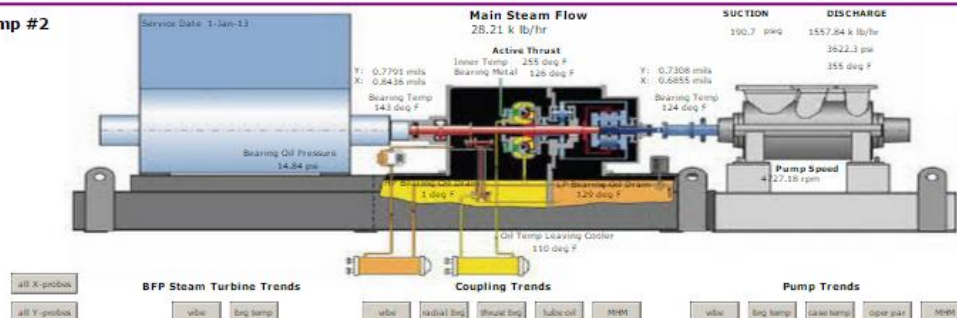
Generator

1 MW status

Pump #1



Pump #2



3/25/2015 7:31:06 AM



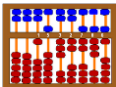







1d

Now

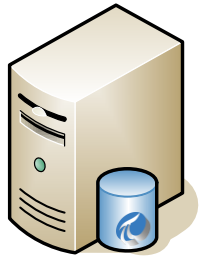
3/26/2015 7:31:06 AM



Time-series data is different – but not hard

- Sensor - data sampling, data rates, data compression... 
- Shaping the data (list, hierarchical...) - data structure, data dictionary... 
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT... streaming data 
- Framing the data – Start/End of a time window...  PI CoreSight
 PI ProcessBook
 PI DataLink
 PI WebParts
- Out-of-the-box - visualization, display tricks, Excel, notifications, smart device... 
- **External** - business intelligence, data mining, machine learning, Matlab, R, Big Data...

PI Integrator



Cleanse

Augment

Shape

Transmit



OSIsoft®

System of Record

- Guaranteed Delivery & Storage
- Full Fidelity of Sensor
- Optimized for Real-Time
- Backup/Restore
- HA
- Security



Analytics Packages

- Designed to Analyze Large Sets
- Expects that the Data Exists
- Problem Defines Data Shape
- Typically Evenly Spaced in Time



Pre-release

The screenshot displays the OSIssoft CAST web interface. The top navigation bar includes buttons for 'Select Data', 'Publish', 'Status', 'Load', 'Save', and 'Refresh', along with a 'Disconnect' button. The left sidebar shows a tree view of assets, with 'USA' expanded to show 'AL', which is further expanded to show 'ALABASTER'. The 'AlabasterMontevalloRoad' asset is selected and highlighted. Below this, a list of attributes for the selected asset is shown, including City, State, Building Code, Region, and various consumption metrics. The main content area is titled 'Target Configuration' and features a dropdown menu with options: File, SQL (Integrated Security), SQL (SQL Authentication), Microsoft Azure Table Storage, Hadoop HBase (selected), Hadoop HDFS, SAP HANA, and SAP Event Stream Processor. To the right of the dropdown, there is a 'Summary' section with the heading 'Shape and Matches' and a bullet point stating 'There are 654 matching instances'. Below this, the 'Timeframe and Interval (Edit)' section shows 'Your Start Time is t-7d.', 'Your End Time is t.', and 'Your Time Interval gets an interpolated measurement every 1h.'. At the bottom right, there is a large blue 'Publish' button. A large 'Pre-release' watermark is overlaid diagonally across the center of the image.





Business Intelligence / Visual Analytics

Date

- 09/06/2014
- 09/07/2014
- 09/08/2014
- 09/09/2014
- 09/10/2014

Hour

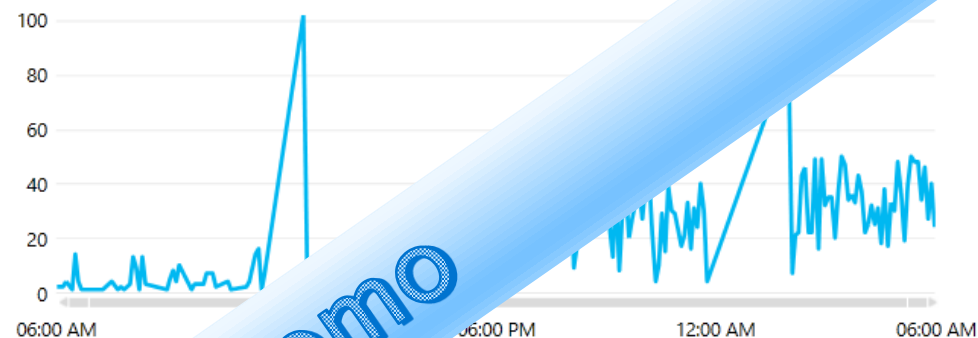
- 01
- 02
- 03
- 04
- 05
- 06
- 07
- 08

Cycle

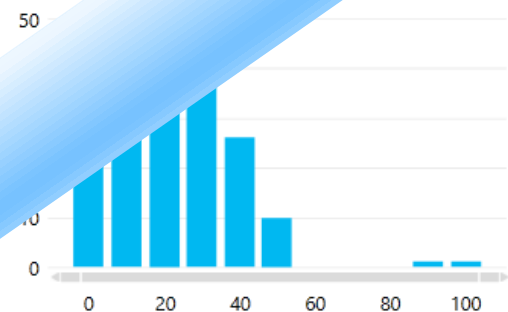
Step 1
Step 2
Str

Cycles - Duration Statistics

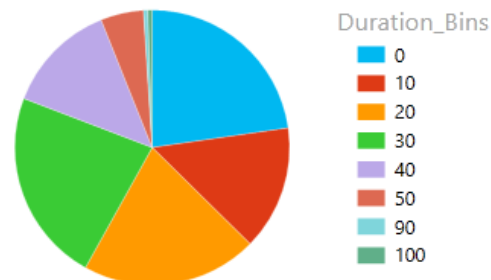
Duration_Seconds by StartTime



Count of Duration_Seconds by Duration_Bins



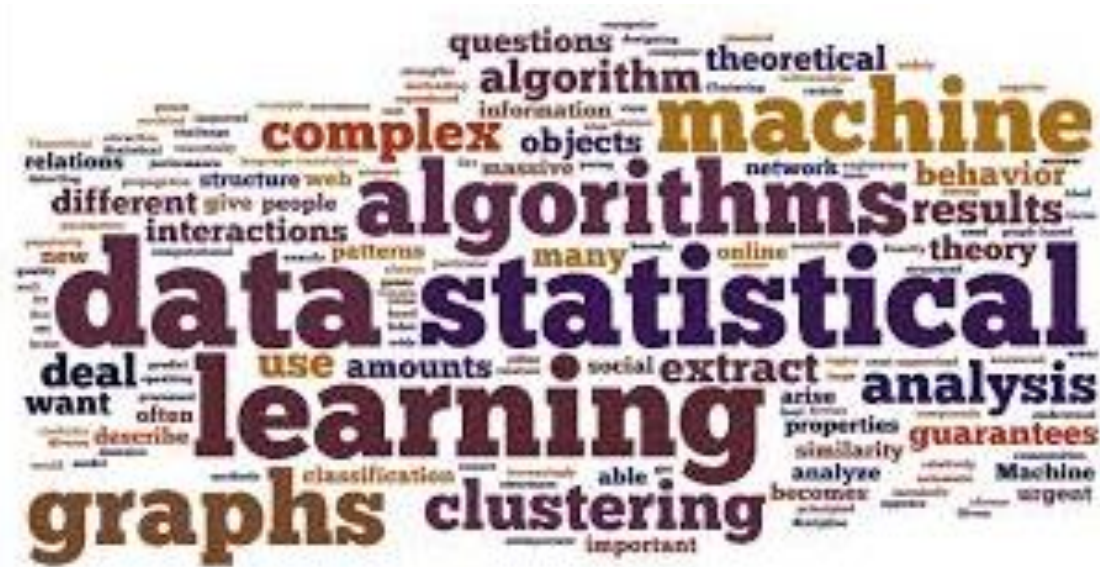
Count of Duration_Seconds by Duration_Bins



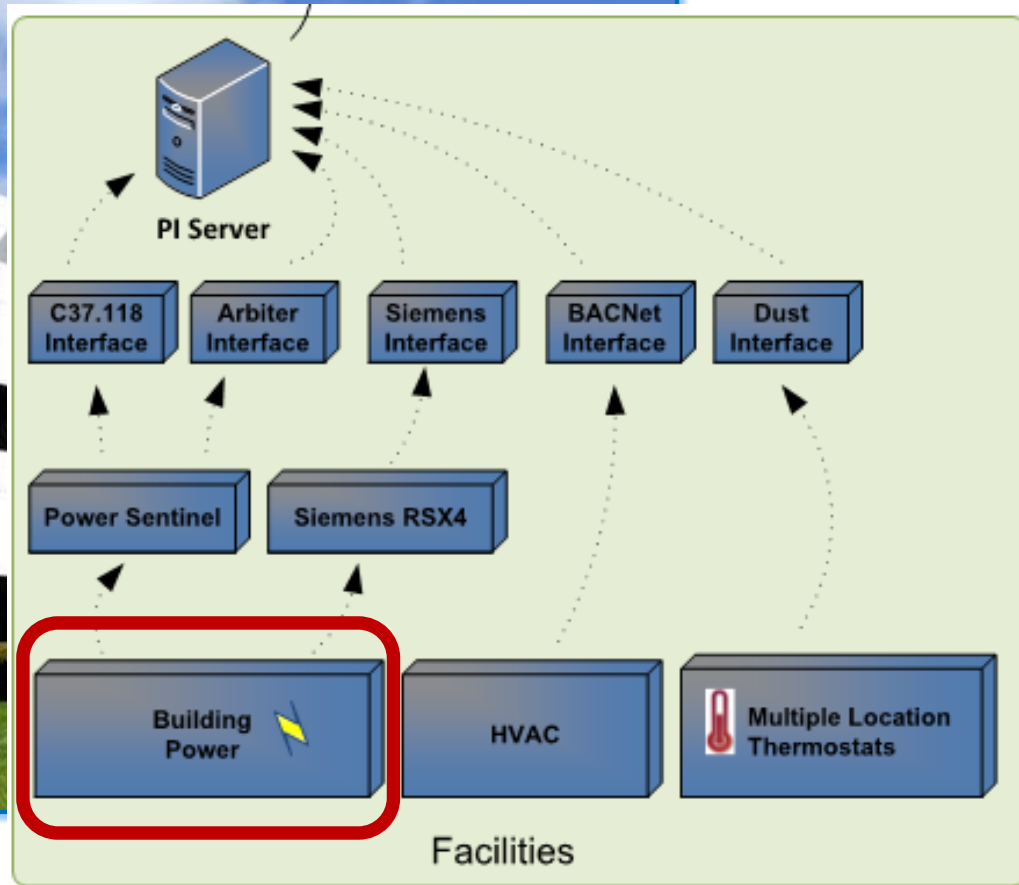
Duration_Bins

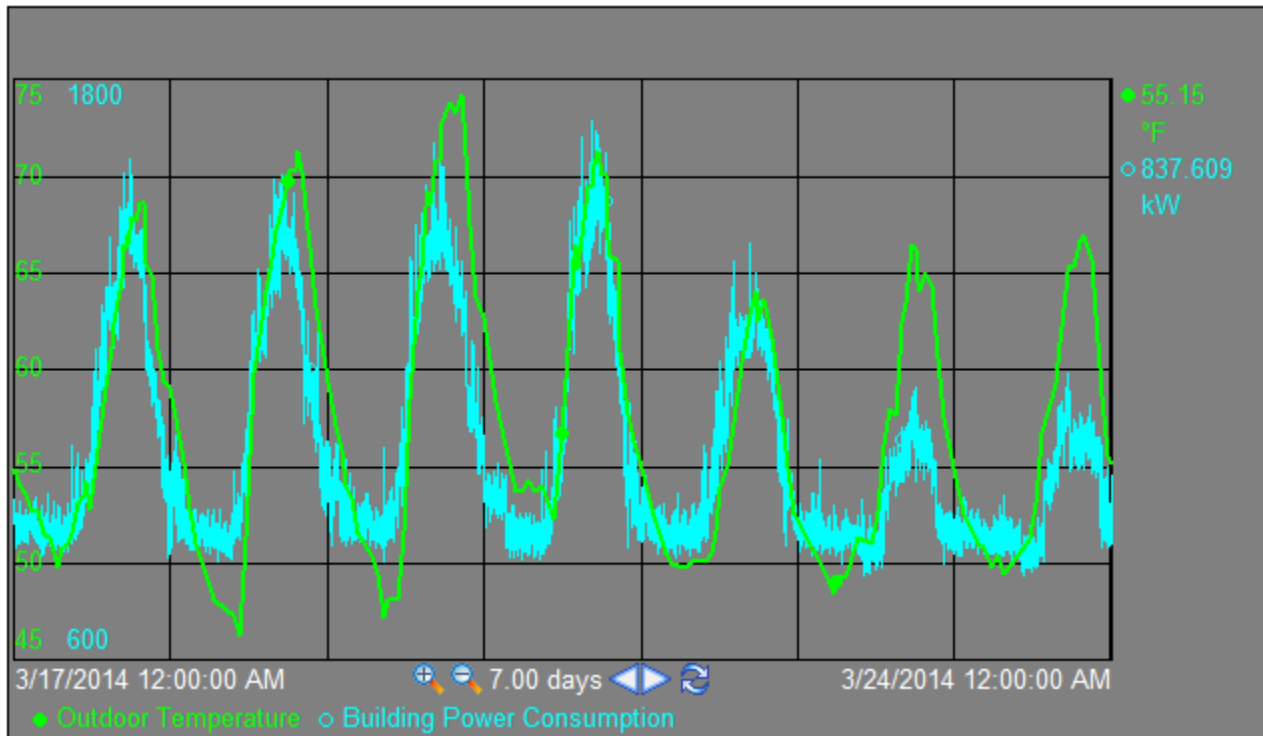
- 0
- 10
- 20
- 30
- 40
- 50
- 90
- 100



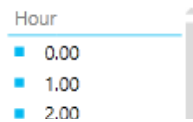
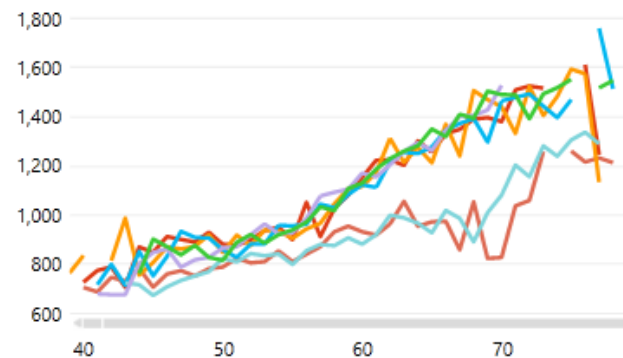


Machine Learning and Predictive Analytics

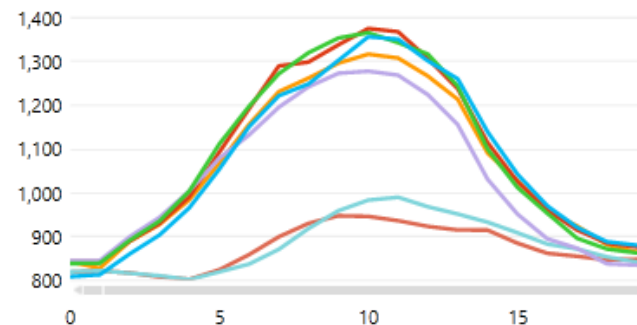




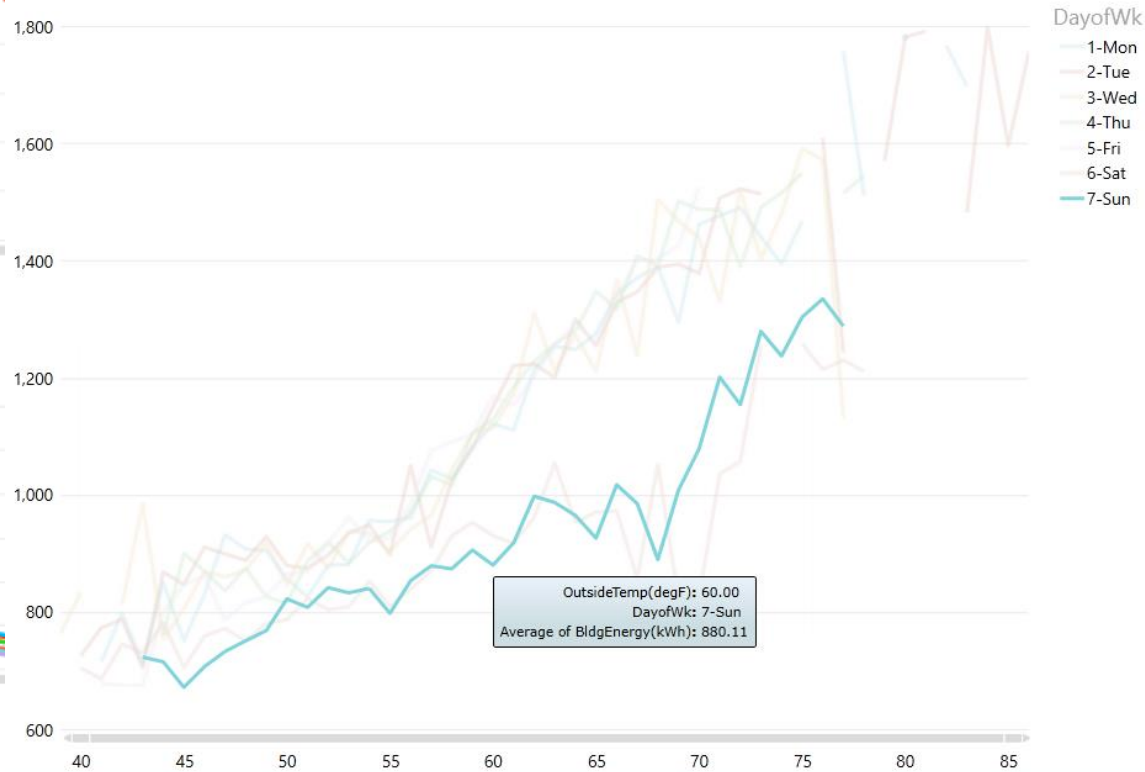
Average of BldgEnergy(kWh) by OutsideTemp(degF), and DayofWk



Average of BldgEnergy(kWh) by Hour, and DayofWk



Average of BldgEnergy(kWh) by OutsideTemp(degF), and DayofWk



BldgPower.csv



Boosted Decision Tree Regress...



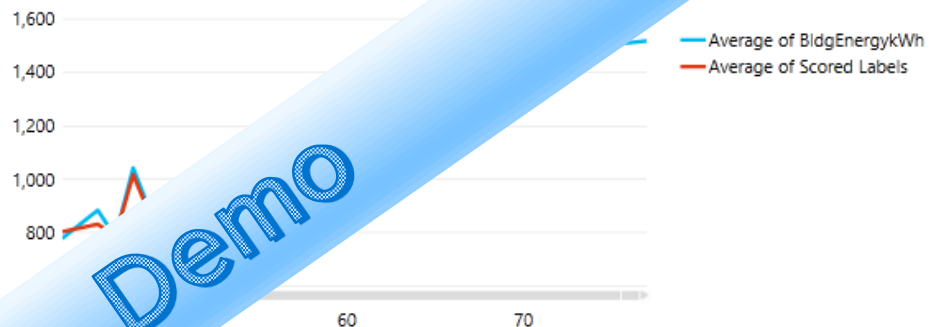
Train Model



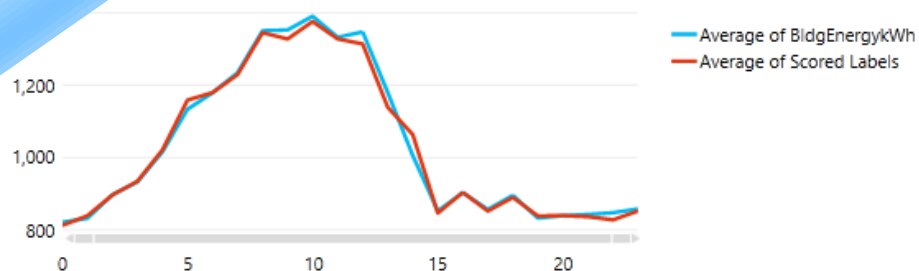
Score Model

Evaluate Model

Average of BldgEnergykWh, and Average of Scored Labels by Hour



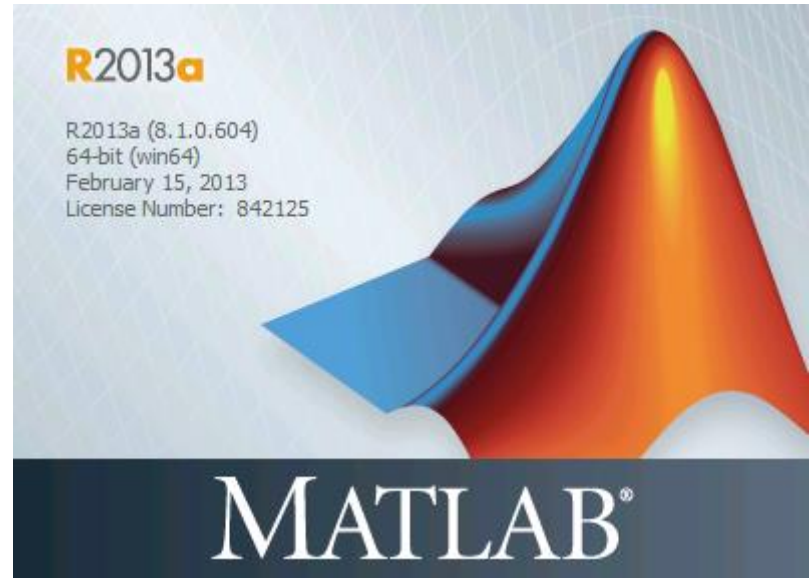
Average of BldgEnergykWh, and Average of Scored Labels by Hour



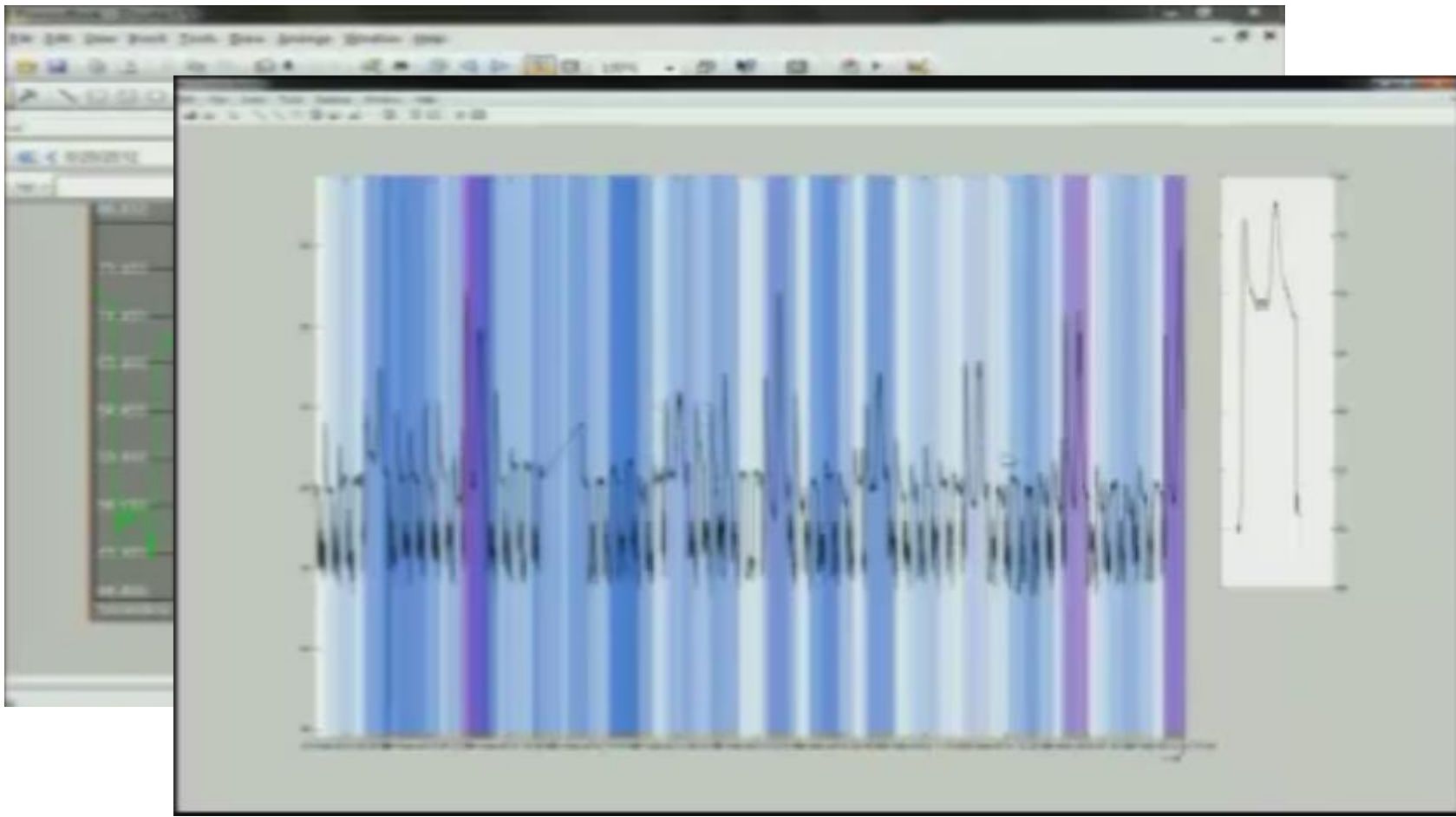
DayofWk

- 1-Mon
- 2-Tue
- 3-Wed
- 4-Thu
- 5-Fri
- 6-Sat
- 7-Sun

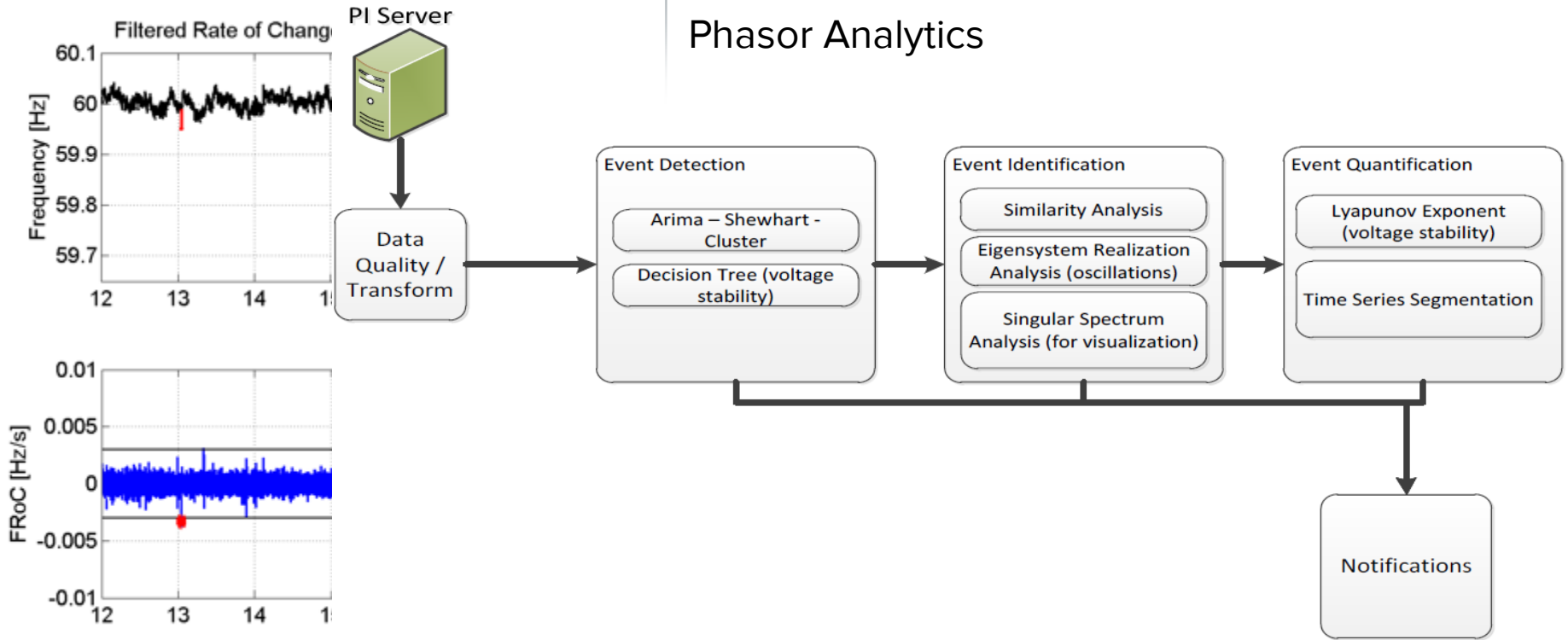




PI System with MATLAB...



Phasor Analytics



http://www.osisoft.com/resources/presentations/presentation_abstracts/2014_-_Industry_Session_-_Transmission_and_Distribution_Smart_Grid/TD140050_Real-time_Event_Detection_of_Microgrid_Dynamics.aspx

NASPI North American
SynchroPhasor Initiative

<https://www.naspi.org/File.aspx?fileID=1343>

Your PI Infrastructure is Essential for Big Data



Insight



Time Series



Internet of Things

Big



Tweet



Merck Optimizes Manufacturing With Big Data Analytics

Pharmaceutical firm uses Hadoop to crunch huge amounts of data so it can develop vaccines faster. One of eight profiles of InformationWeek Elite 100 Business Innovation Award winners.



Llado (left) and Megaro used cloud-based Hadoop computing to speed up the analysis of vaccine yield rates.

<http://www.informationweek.com/strategic-cio/executive-insights-and-innovation/merck-optimizes-manufacturing-with-big-data-analytics/d/d-id/1127901>

Gopal GopalKrishnan, P.E.

gopal@osisoft.com

Solution Architect

OSIsoft, LLC

Time-series data is different,
but not hard...

Thank You



OSIsoft®