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:

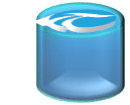
"Visualization tools… are putting business users in control of the analytics tools by making it easy to slice and dice data, define the data exploration needed to address the business issues, and support decision making," writes David Court. Earlier in the article he says "…analytics specialists builds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

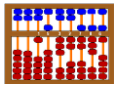
Sensor data, i.e. time-series data from aircraft or ground assets or from manufacturing operations, has its own flavor of bigness – along with its three V's - volume, velocity, and variety to make up the industrial big data. But we have successfully dealt with it across numerous industries – in power 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 da Rows not aligned

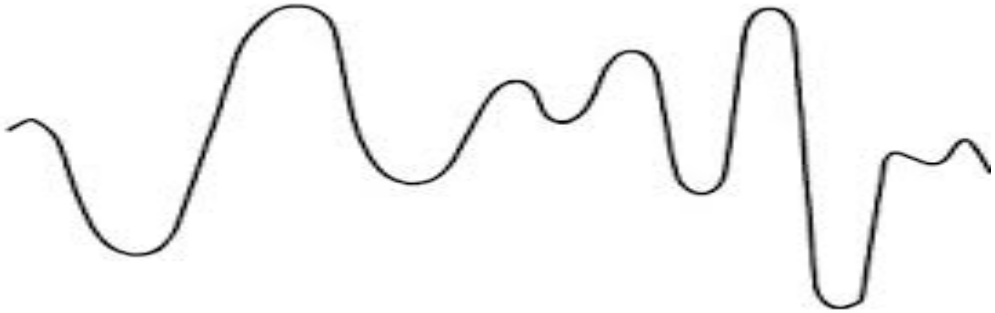
MySensor 3.14 04-16-2015 01:10:27 PM

1429189827 sec POSIX UTC

1429189827.35897 sec

~ 15 microseconds precision

Instruments and sensors – data, data sampling, data rates, data compression…



Dynamic Sampling 10 µHz ~ 50 kHz

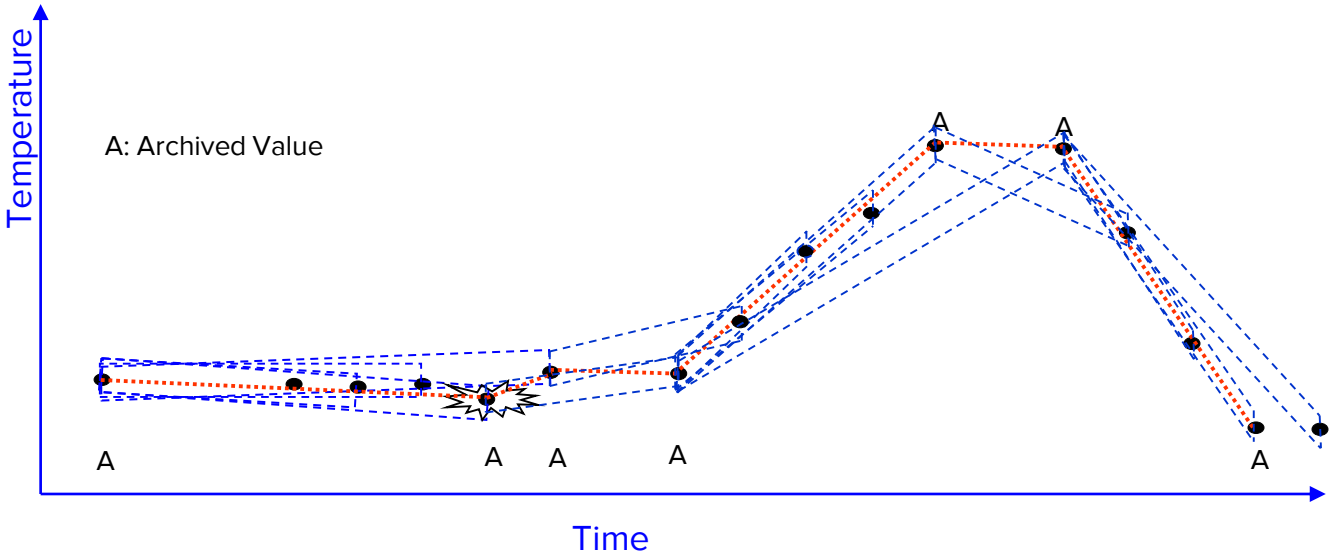
once a day to microseconds

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 Data Archive

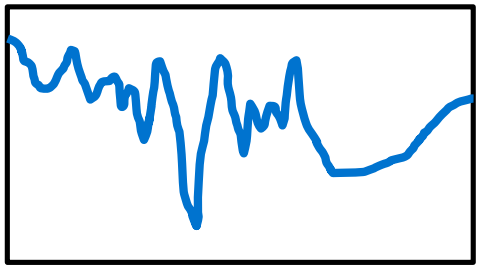
PI AF (Microsoft SQL)

Rotor Speed

GT56.TIC.PV

- Caribou
- Glen Dhu
- Kent Hills
  - WTG 10
  - WTG 11
    - Rotor Speed
    - Wind Direction

WT11.SI.PV



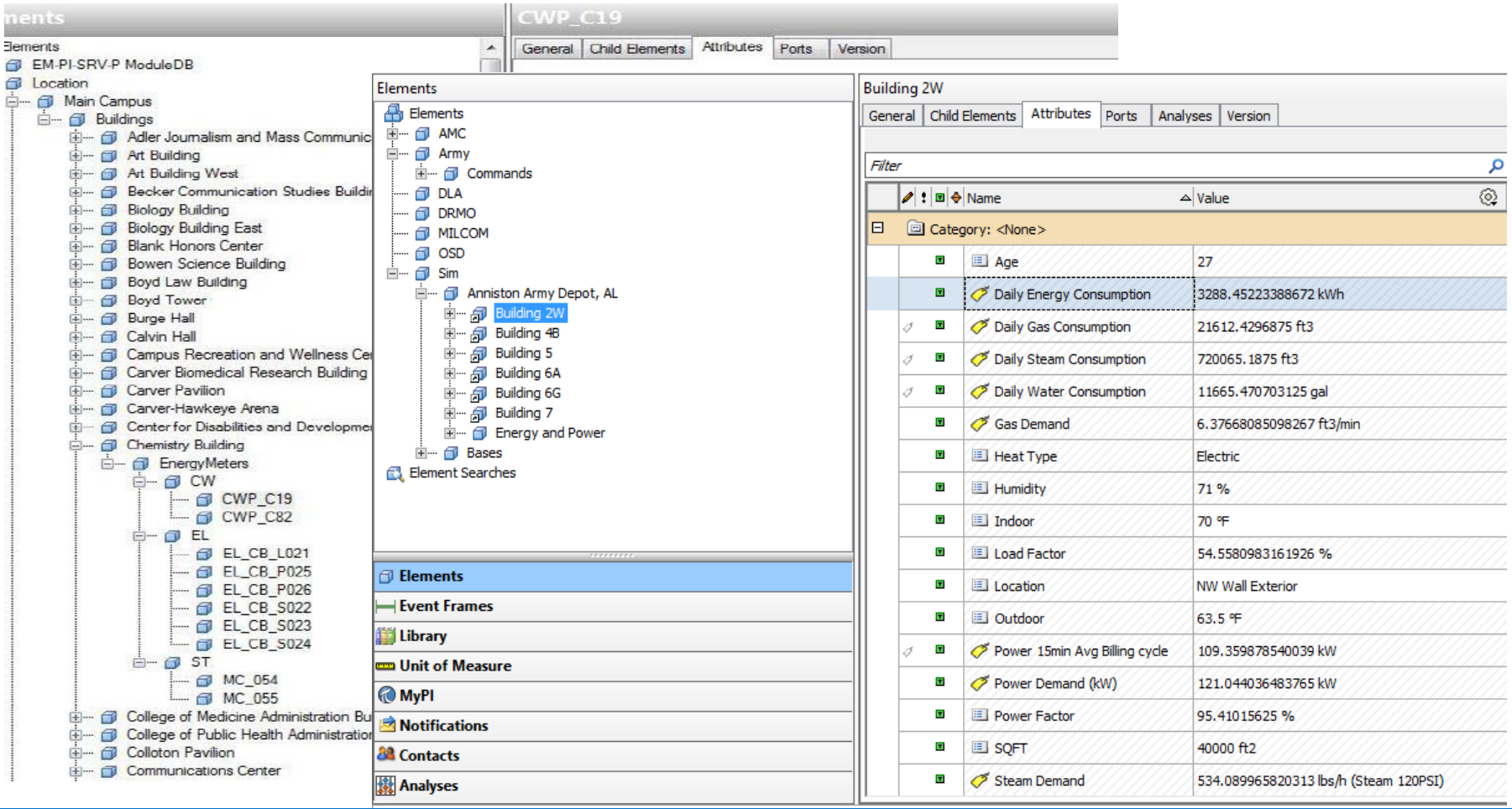
Efficiency

P56.PEF.CALC

- Pump 54
- Pump 55
- Pump 56
  - Efficiency
  - Pressure In
  - Pressure Out

GP23.ATHK8.PV

	A	B
1	Efficiency	74.54%
2		
3		

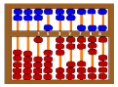


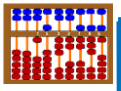
Time-series data is different – but not hard

- 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





Calculation – Rollup

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

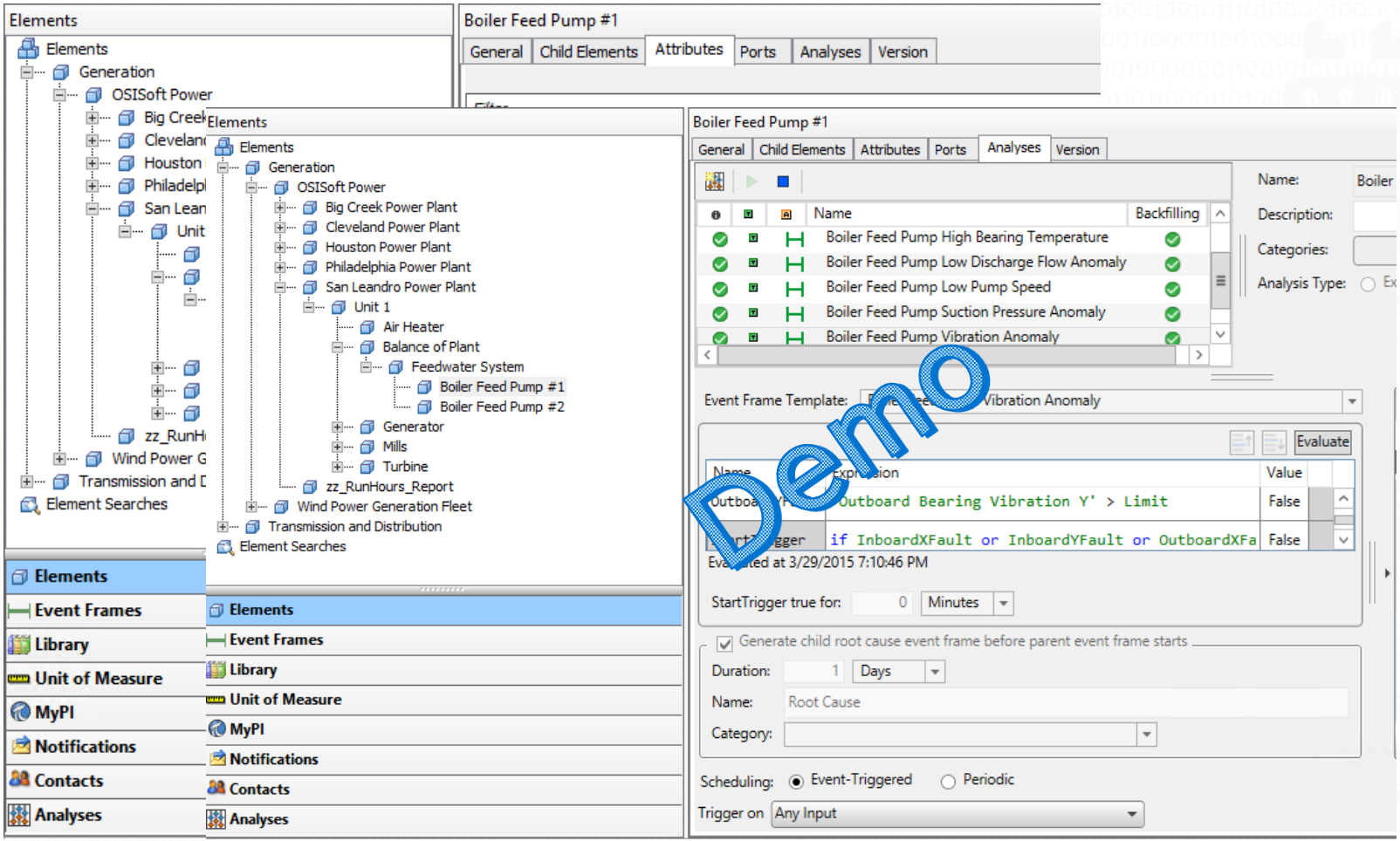
Efficiency

Boiler3

Flow Out

Fuel Flow Rate

Efficiency



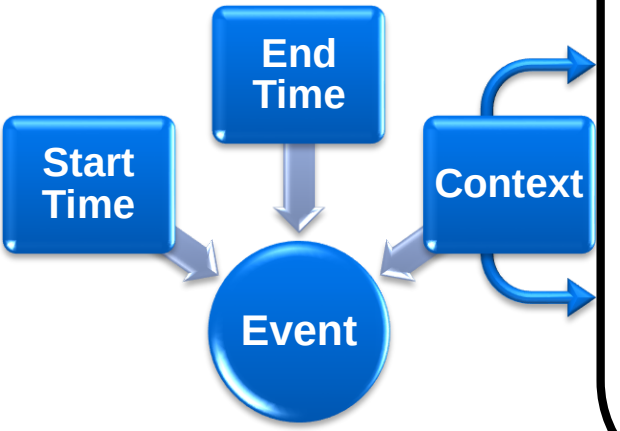
Time-series data is different – but not hard

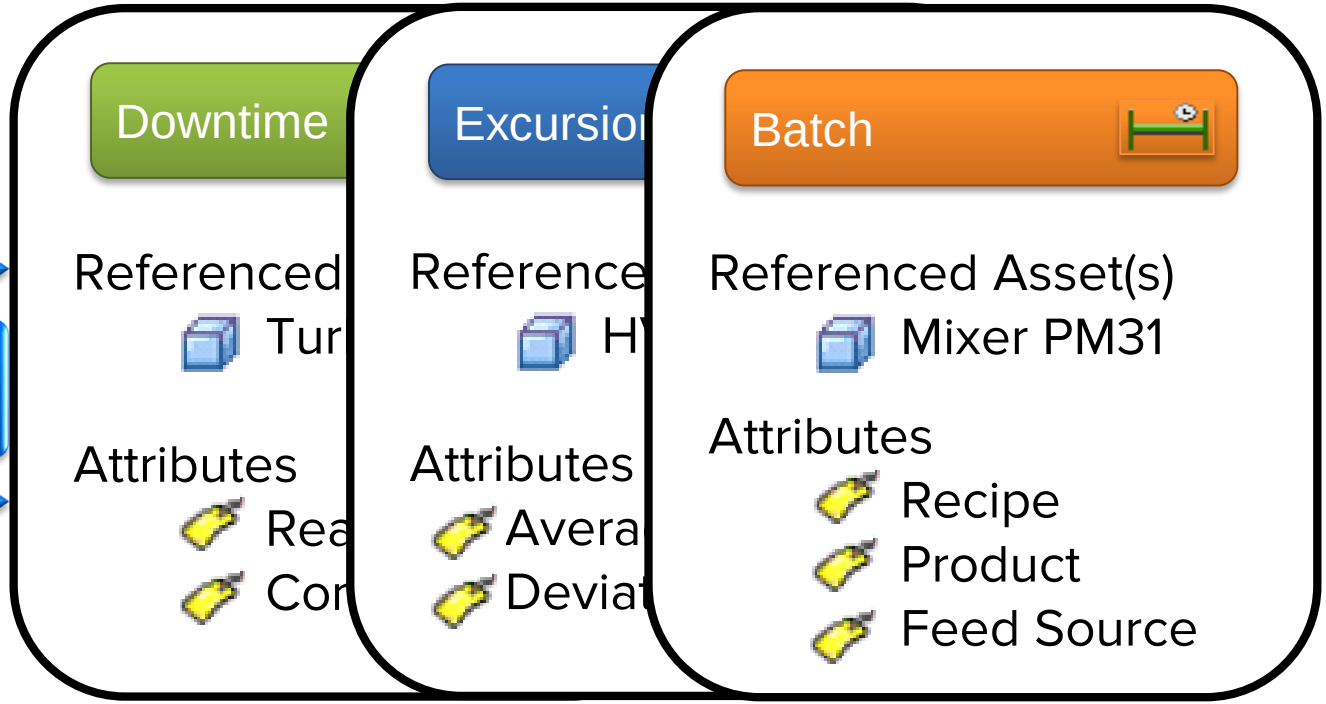
- Sensor - data sampling, data rates, data compression…
- Shaping the data (list, hierarchical…) - data structure, data dictionary…
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT… streaming data
- **Framing the time-series data - Start/End of a time window…**
- Out-of-the-box - visualization,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

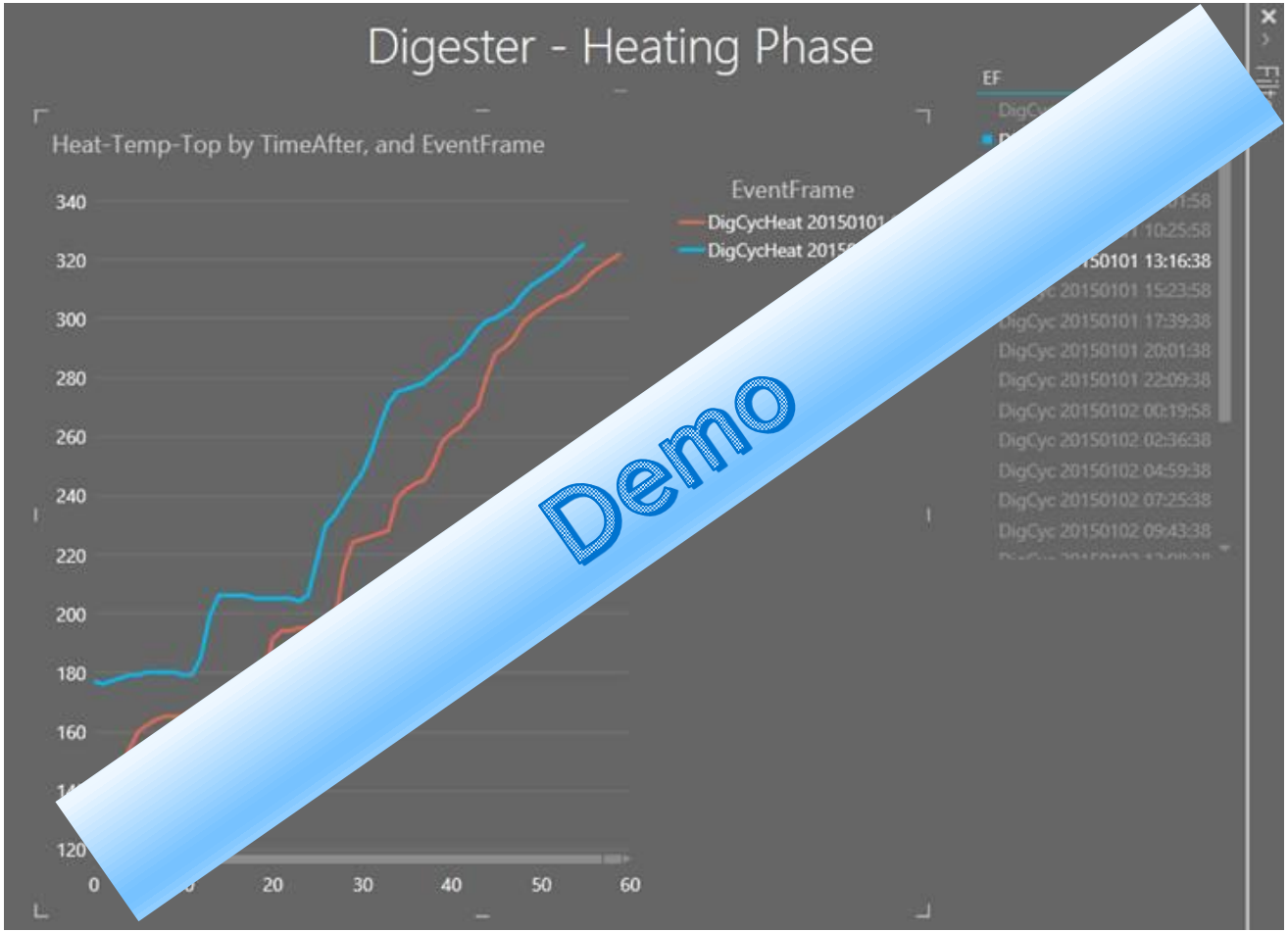
Framing the time-series data - Start/End of a time window…

Define your Events

PI EF (Microsoft SQL)

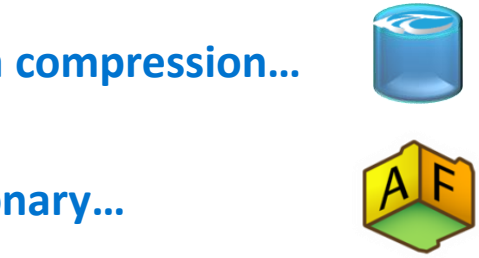






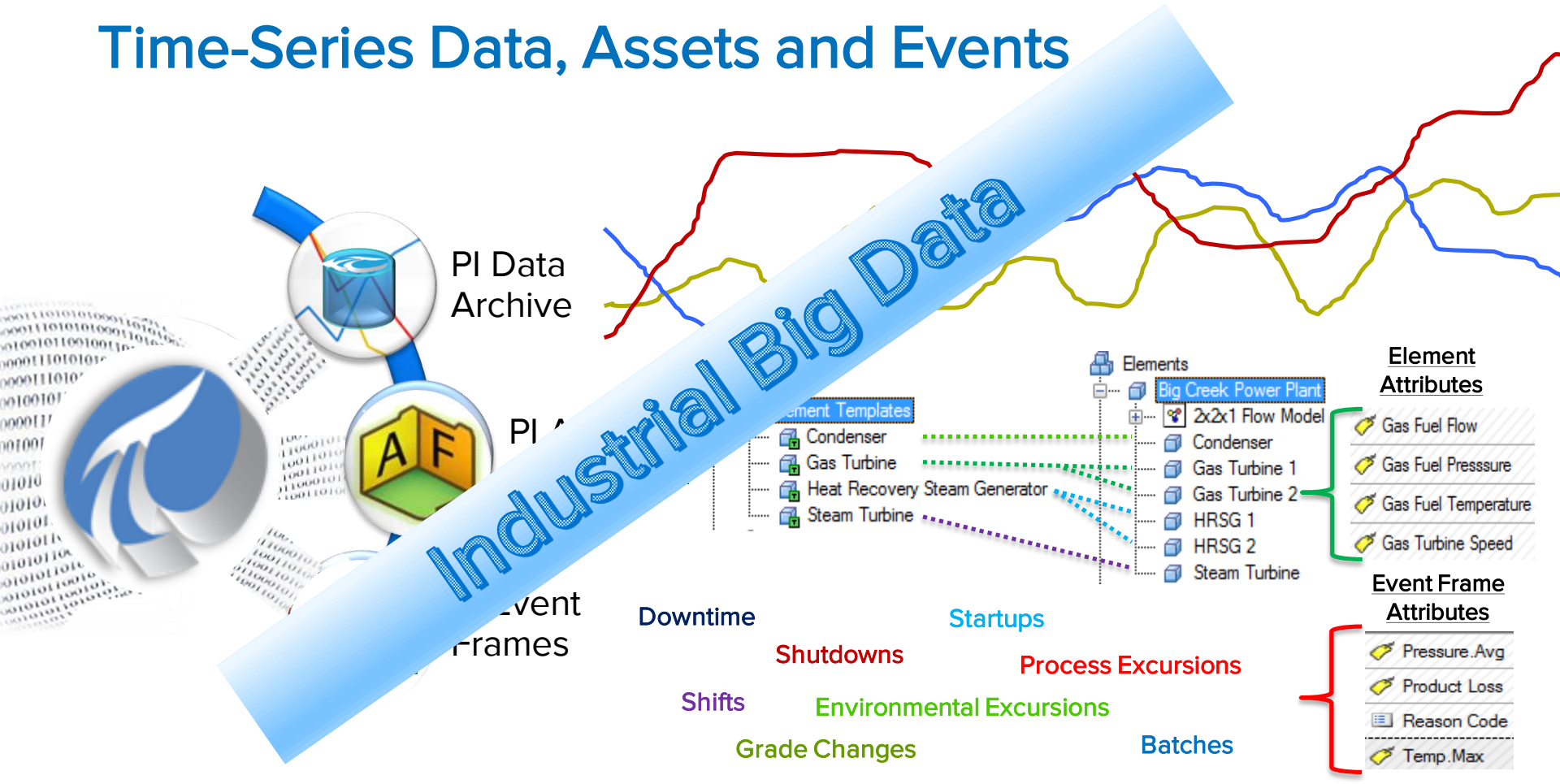
Time-series data is different – but not hard

- **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…

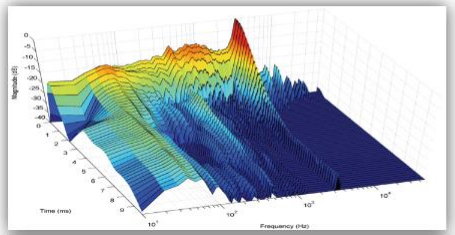
Time-Series Data, Assets and Events



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

430TB

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

Data Center

840TB

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

Automated Metering

1,410 TB

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

Fleet Monitoring

50,460 TB

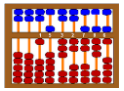
1K assets, 1M points
10 years online
Unique Events: 6,307 Tr
Estimated Data: 50,460TB

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…

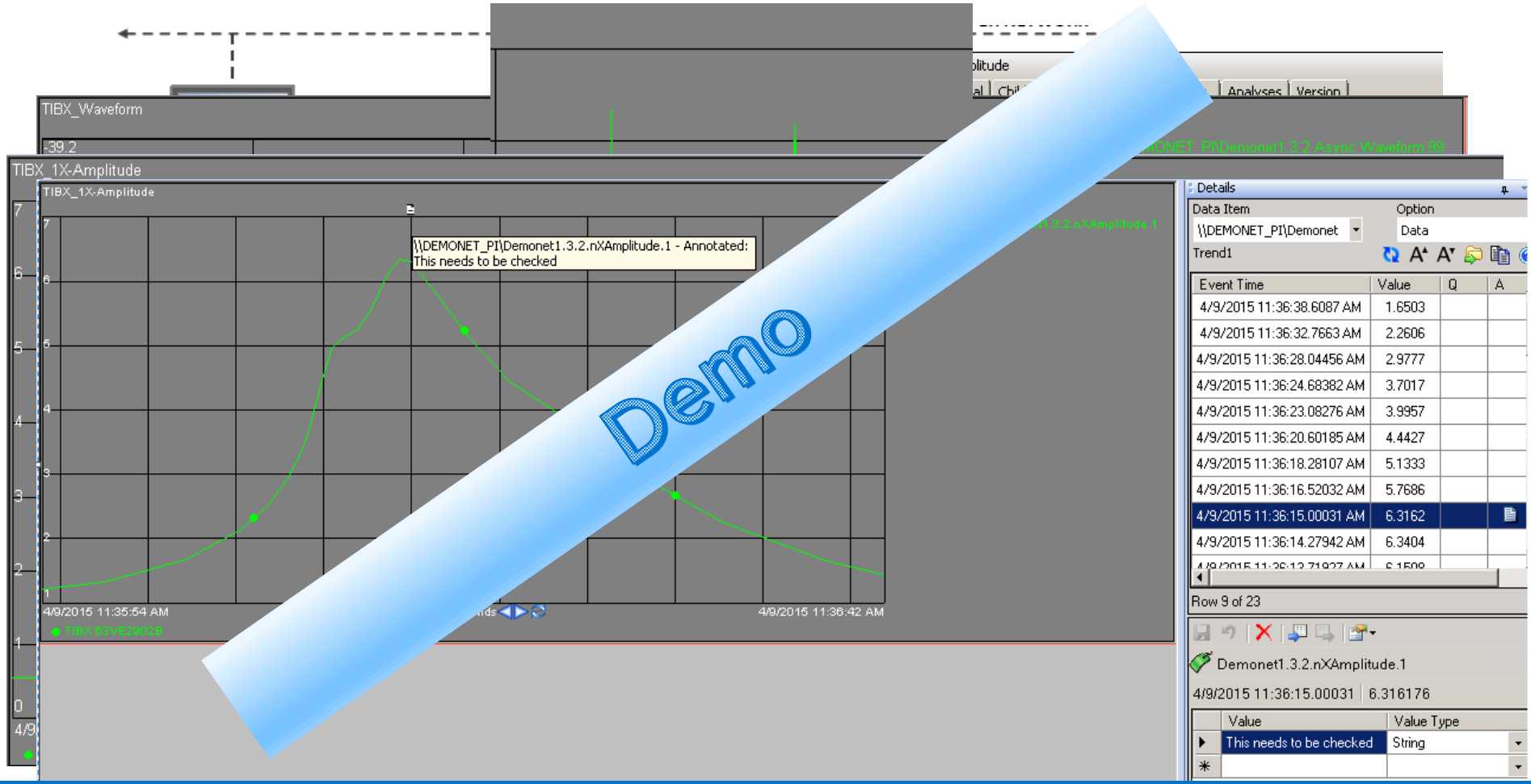
- Calculations - virtual sensor, expressions, roll-ups, SQC, FFT… streaming data



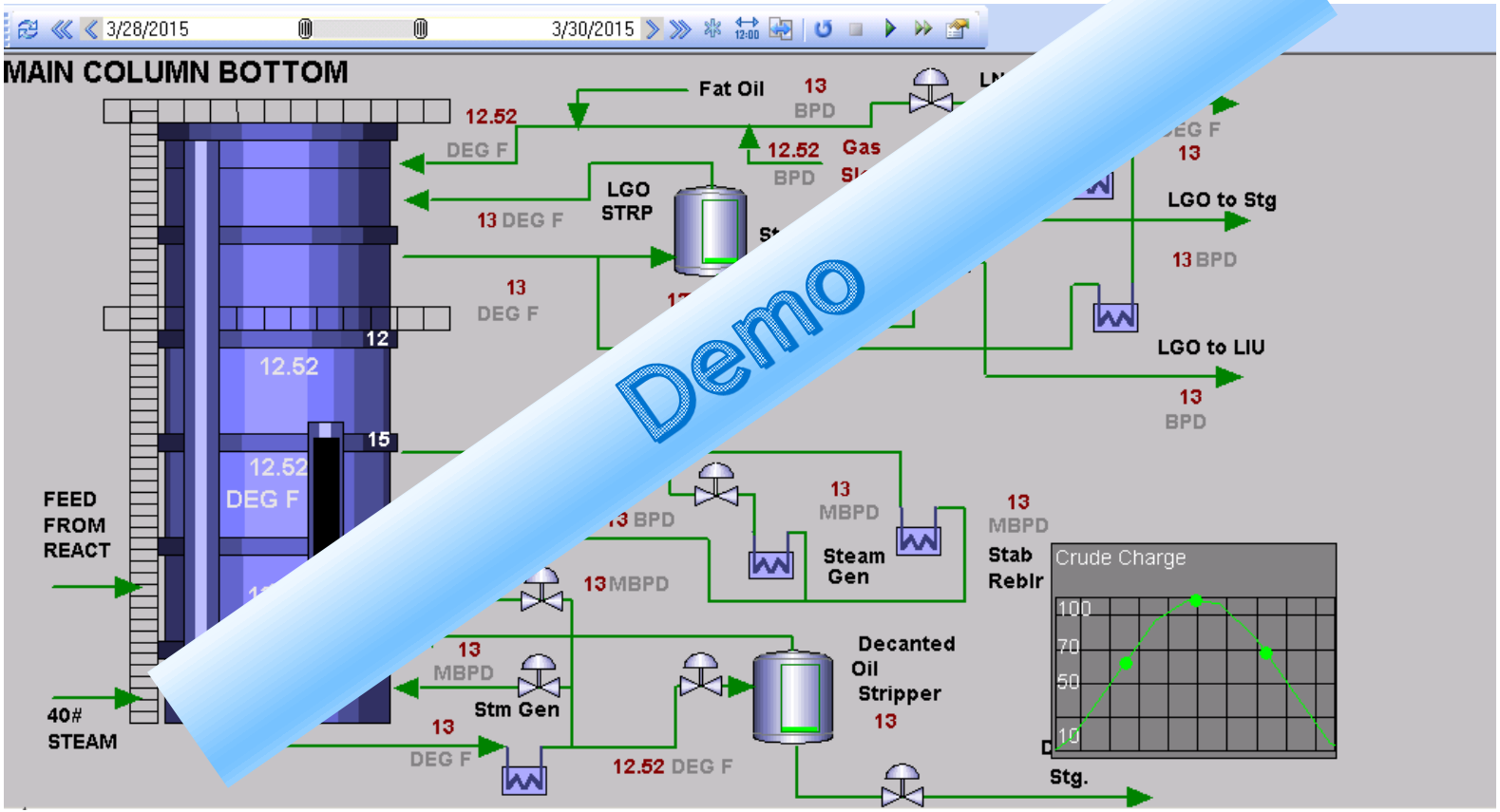
- Framing the data - Start/End of a time window…

- **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…

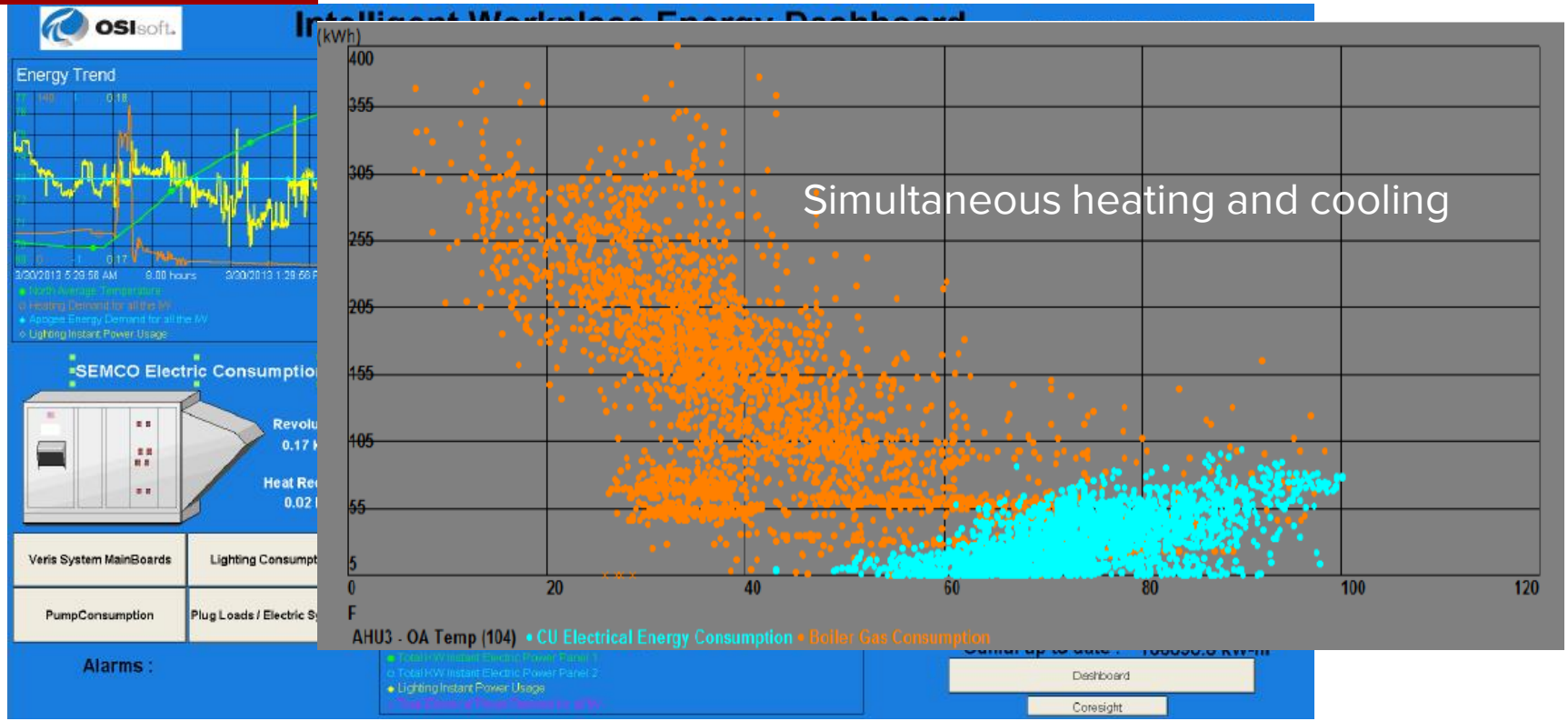


Replay time-series data



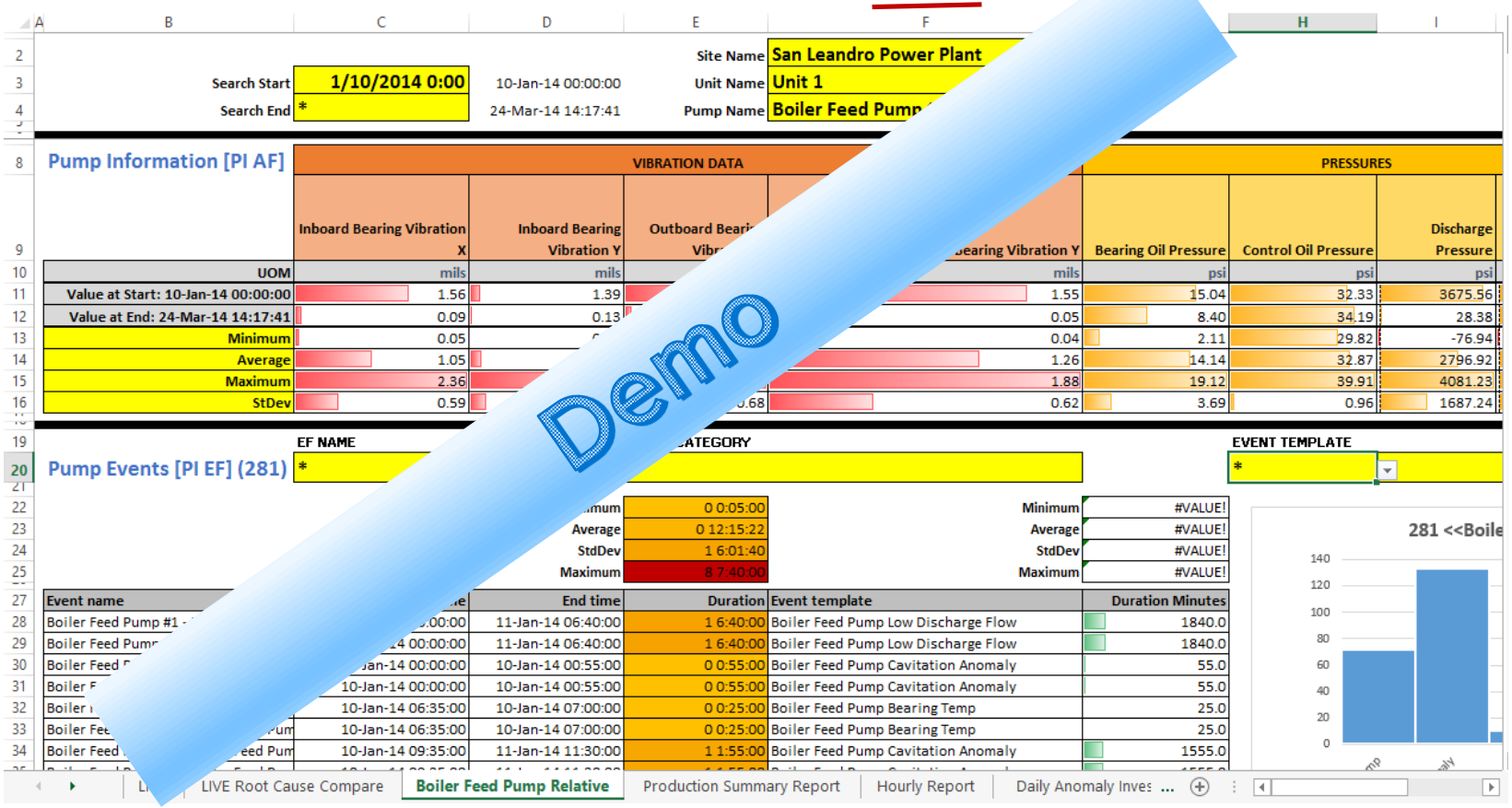
Carnegie Mellon

Facility manager's screen



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

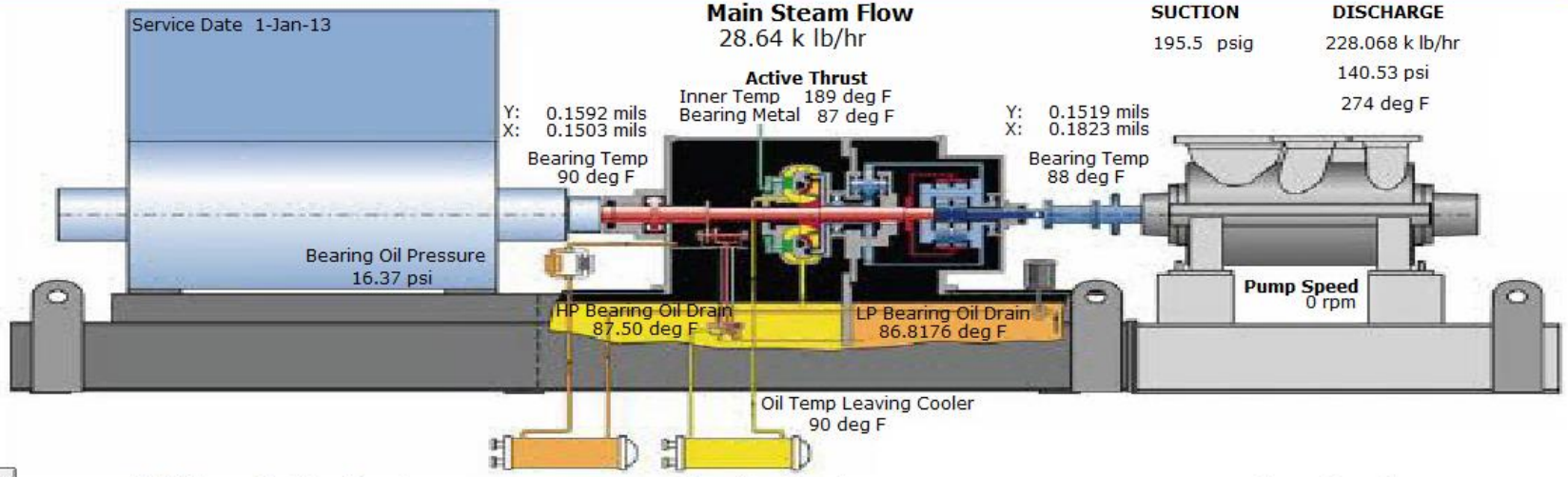
Out-of-the-box - visualization, display tricks, replay, Excel, notifications, smart device…



Boiler Feed Water Pumps

Analysis | Station OvrVw | Generator | 2 MW gross

Pump #1



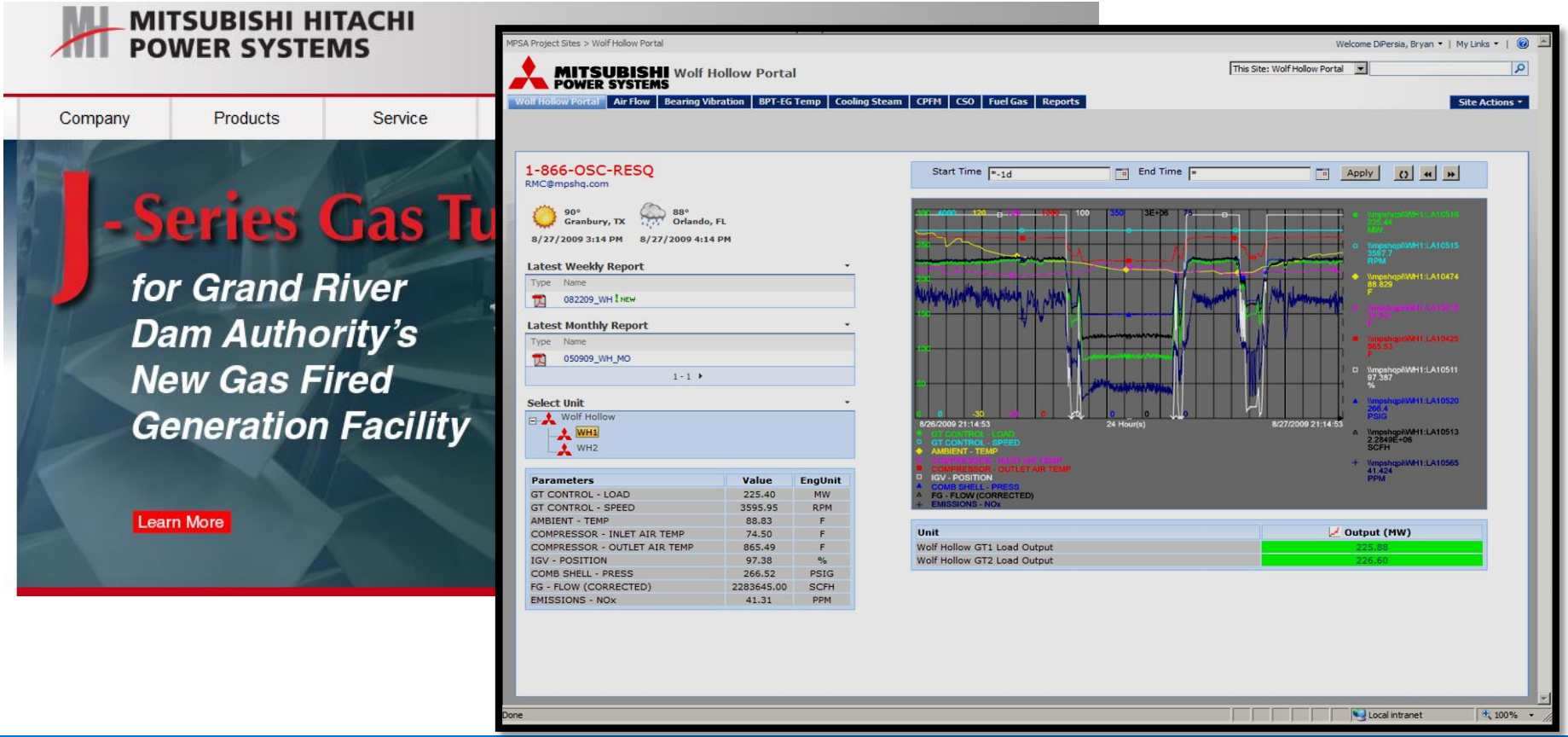
all X-probes
all Y-probes

BFP Steam Turbine Trends: vibe | brg temp

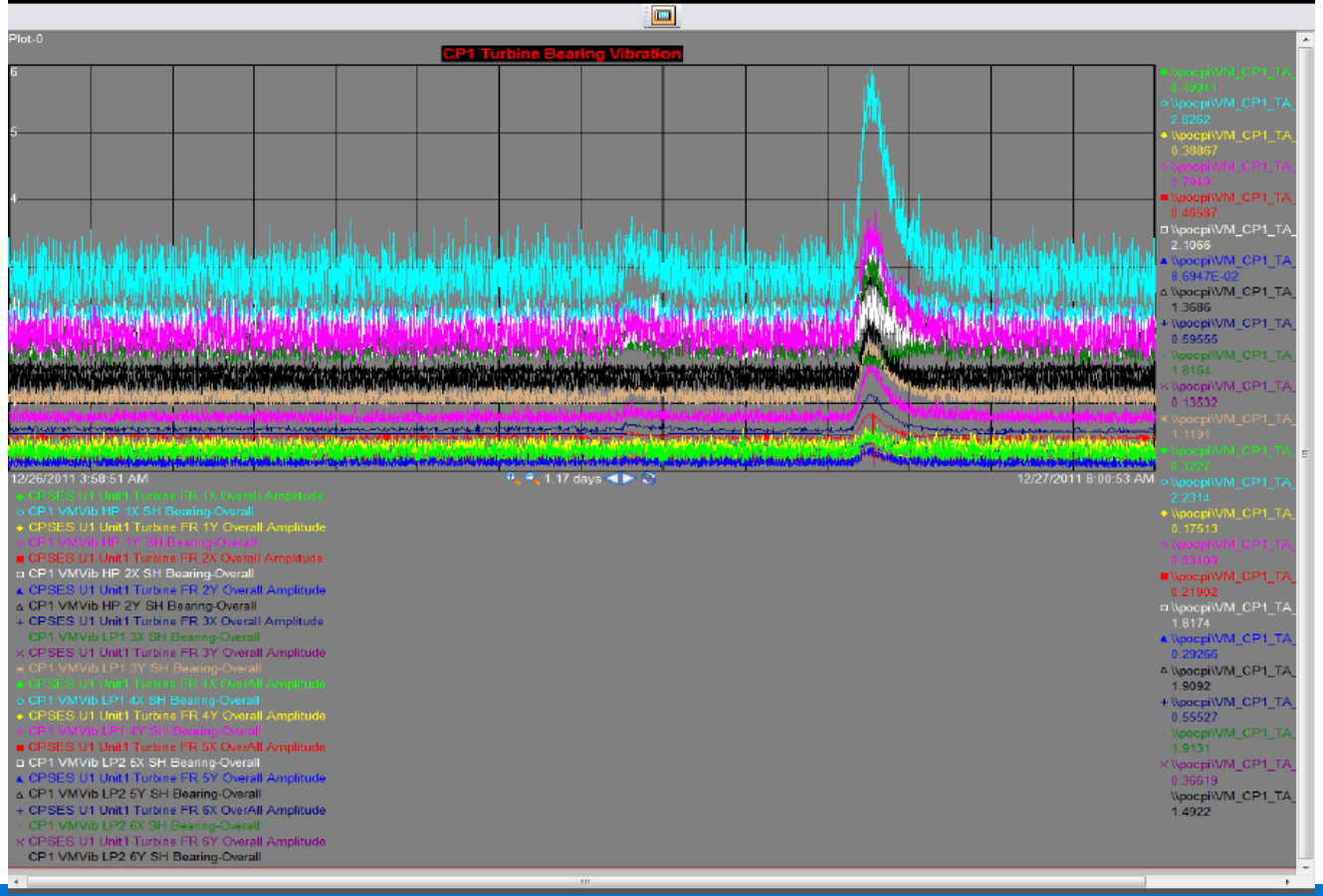
Coupling Trends: vibe | radial brg | thrust brg | lube oil | MHM

Pump Trends: vibe | brg temp | case temp | oper par | MHM

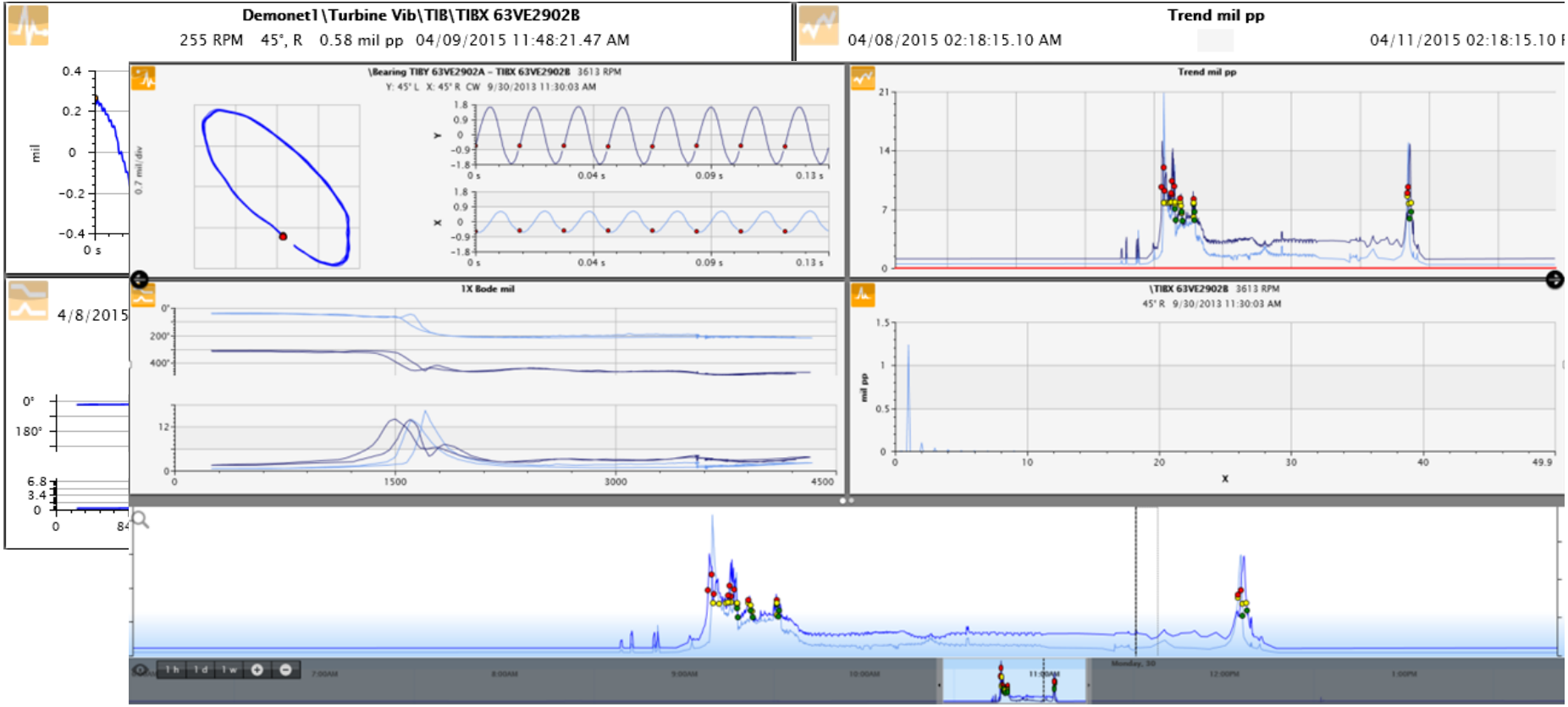
Remote monitoring – several hundred gas turbines



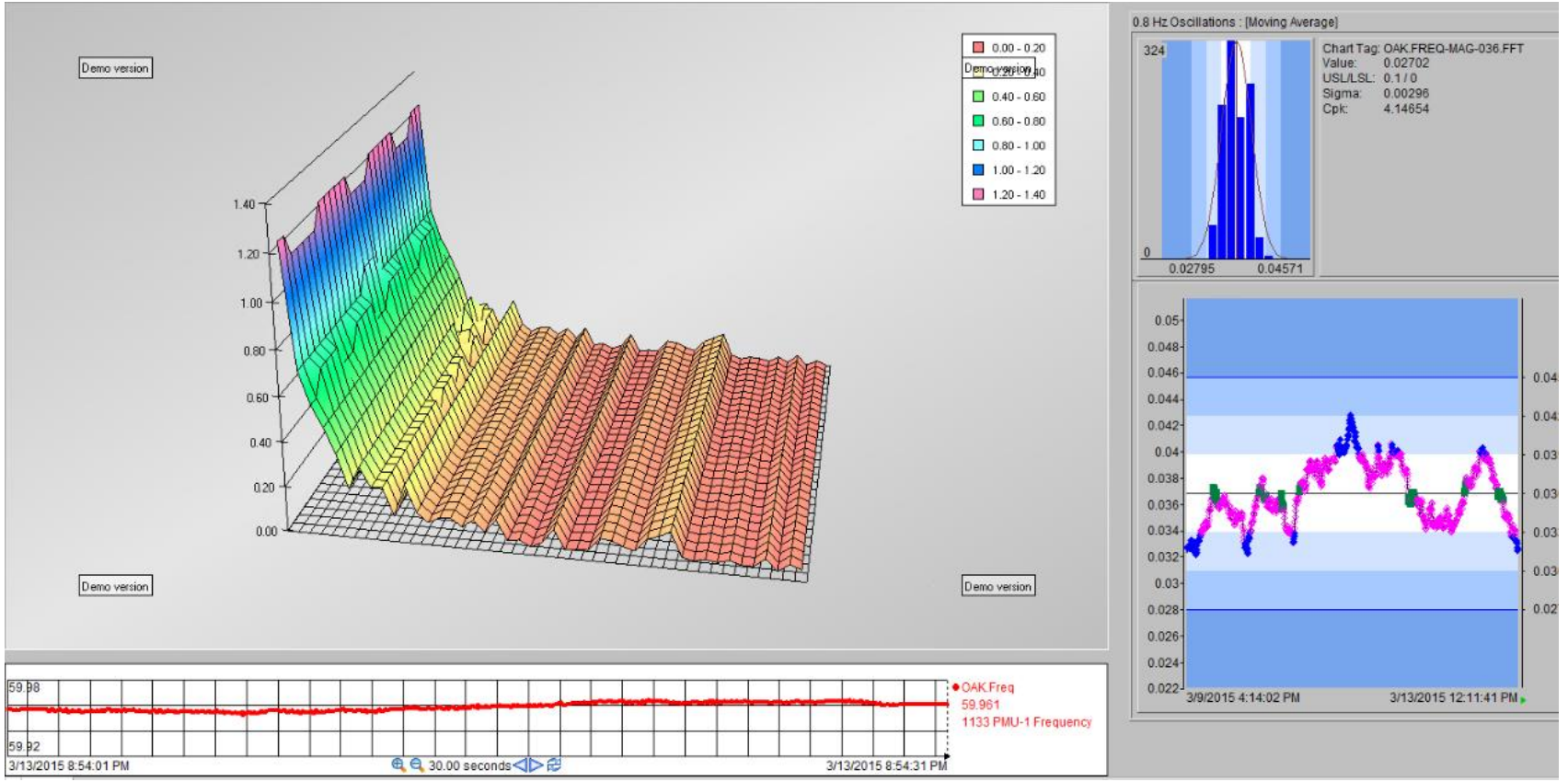
Turbine Vibration Excursion – Comanche Peak Nuclear



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

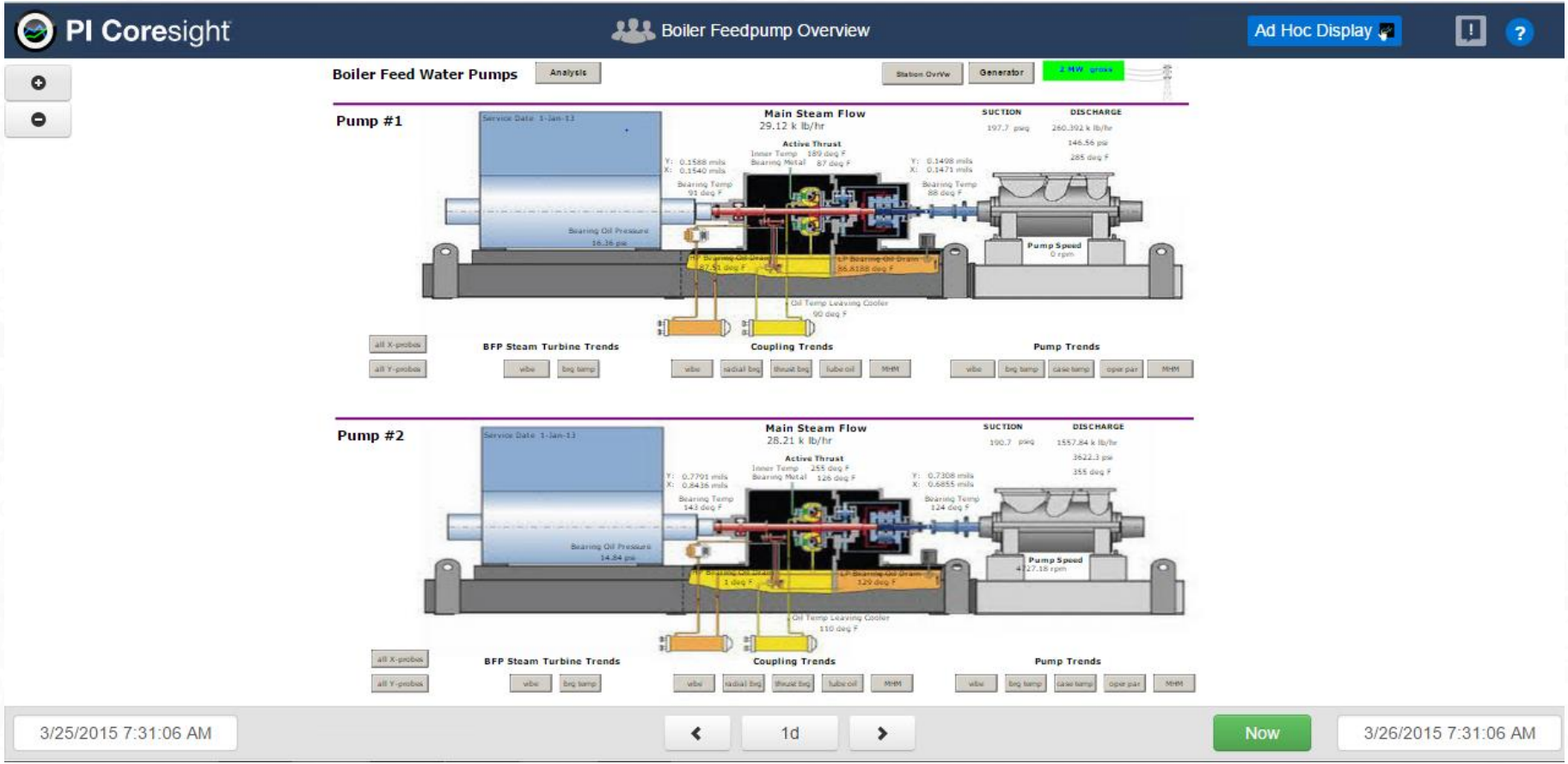
F3 Total Pounds Produced = {Error retrieving result} lb

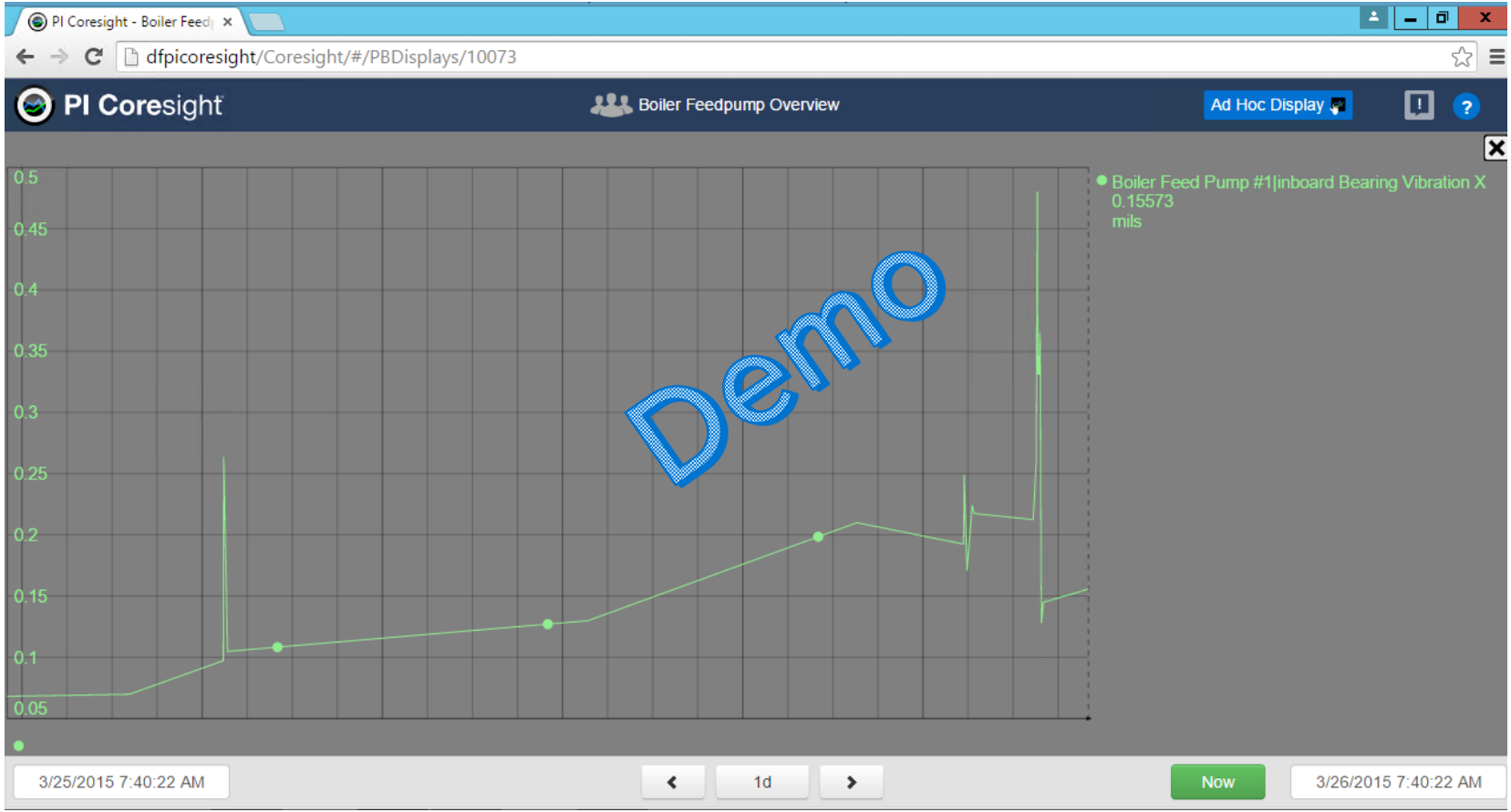
F3 Downtime = 2.166667 h

Cake Wash = 56814.78 US gal

Slurry Feed = 12145.21 US gal

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



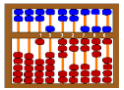


Time-series data is different – but not hard

- 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

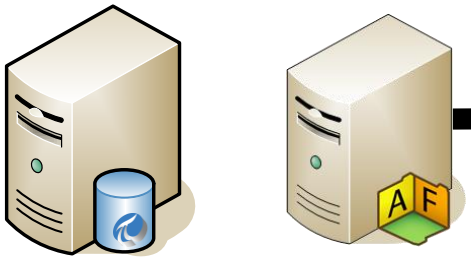


- Framing the data – Start/End of a time window…

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

- **External** - **business intelligence, data mining, machine learning, Matlab, R, Big Data…**

PI Integrator



OSIsoft.

System of Record

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

Cleanse

Augment

Shape

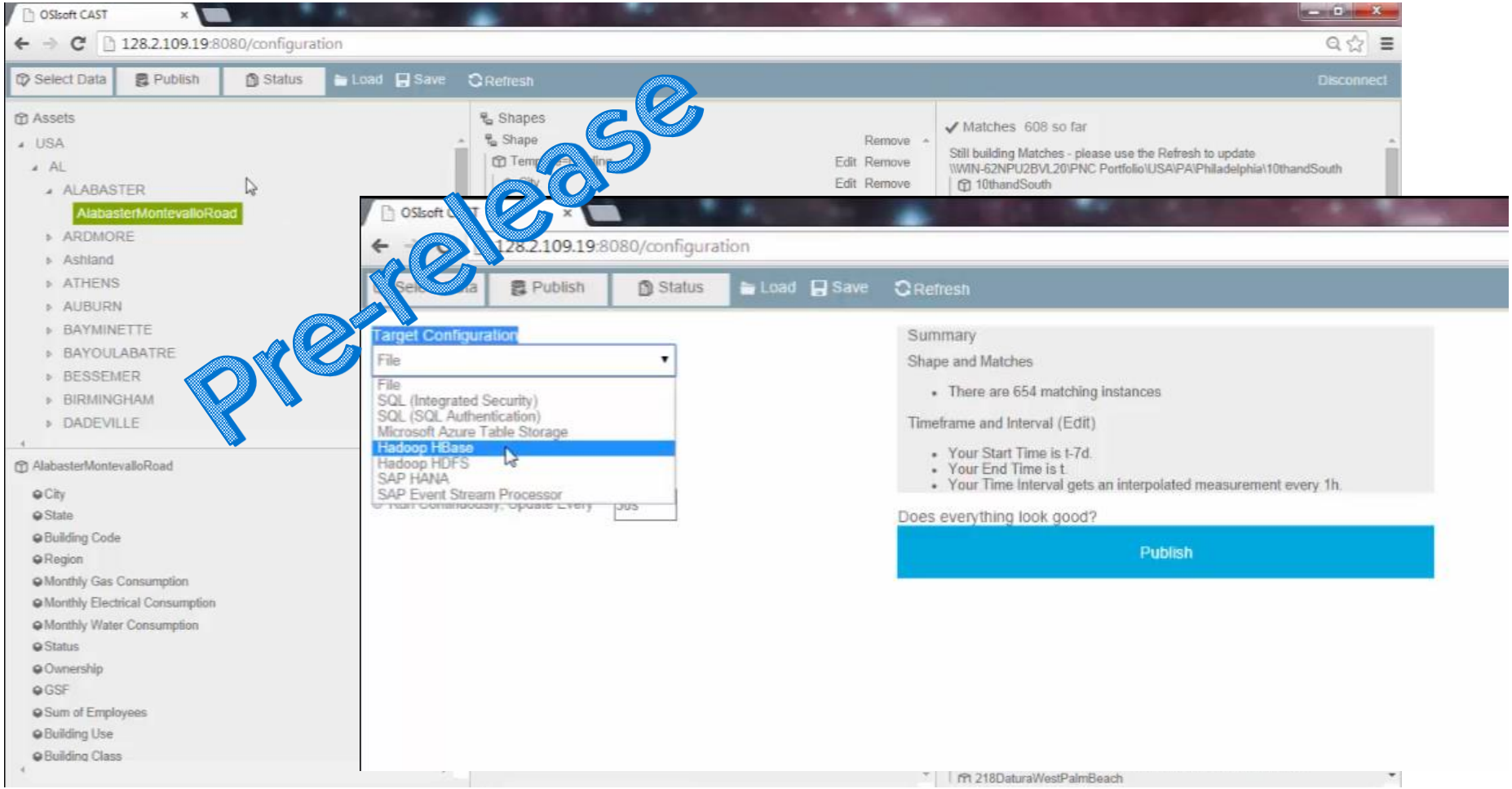
Transmit

Excel tableau Spotfire TIBCO Software COGNOS QlikView

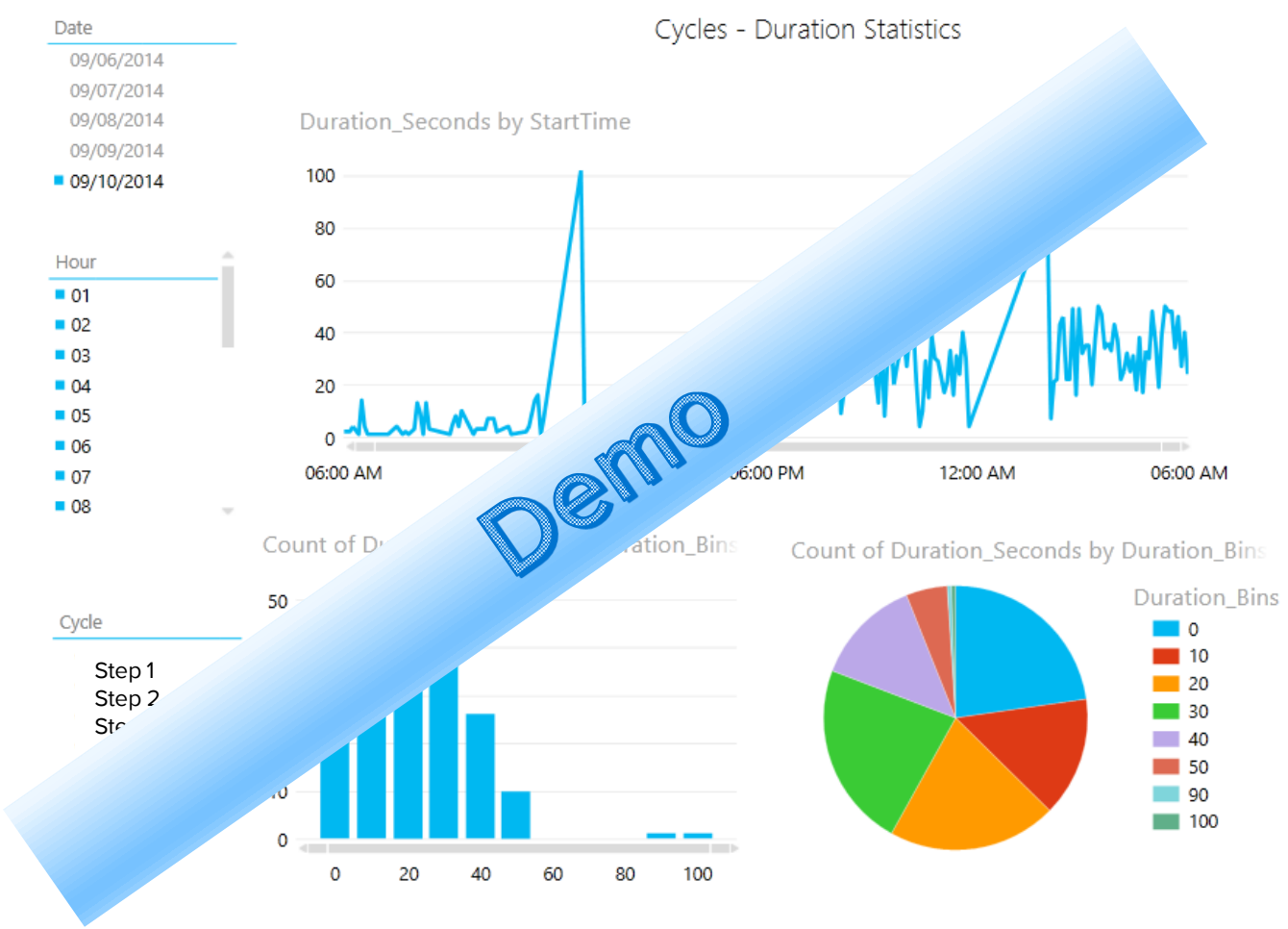
hadoop MapReduce SAS SAP

Analytics Packages

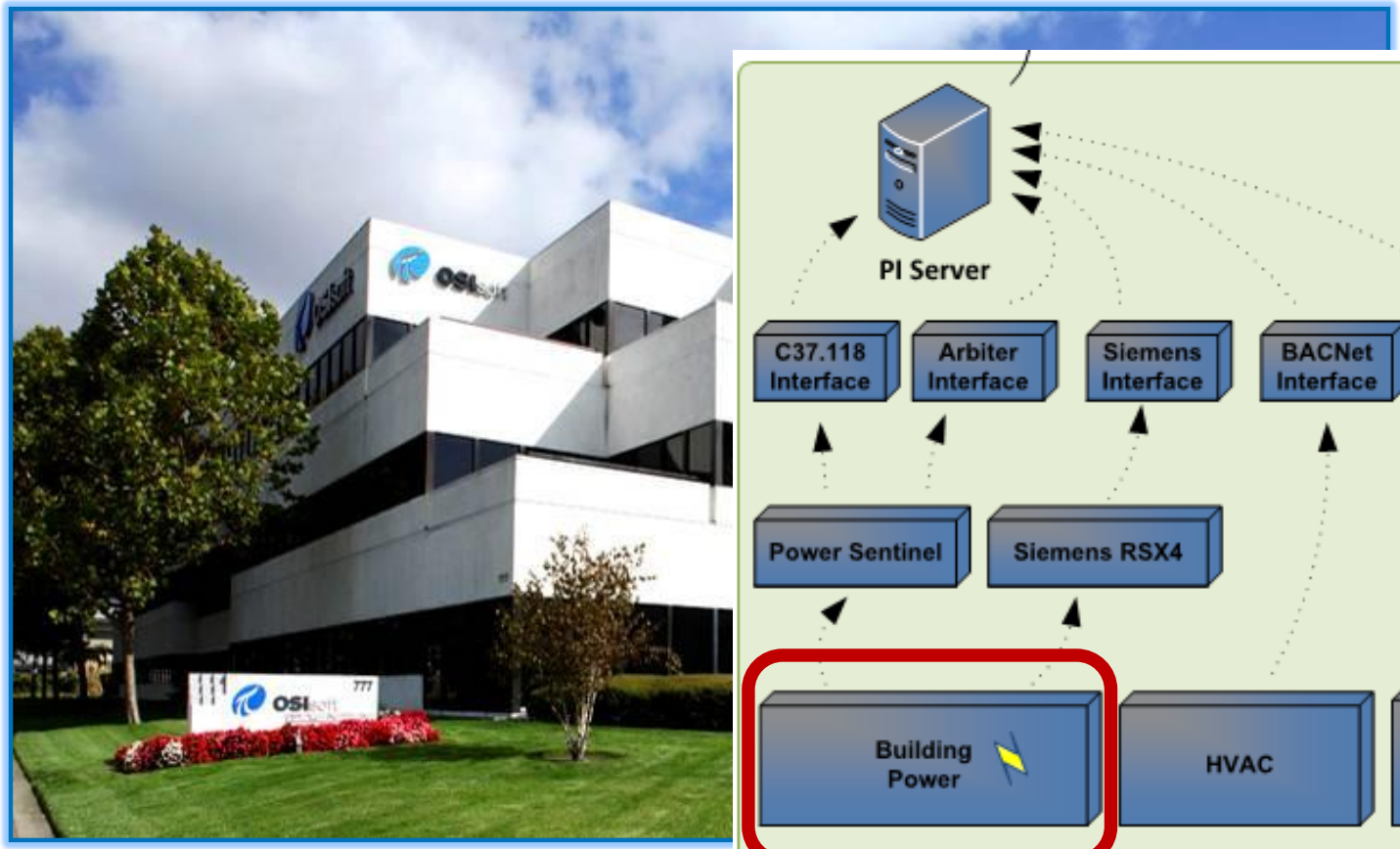
- Designed to Analyze Large Sets
- Expects that the Data Exists
- Problem Defines Data Shape
- Typically Evenly Spaced in Time

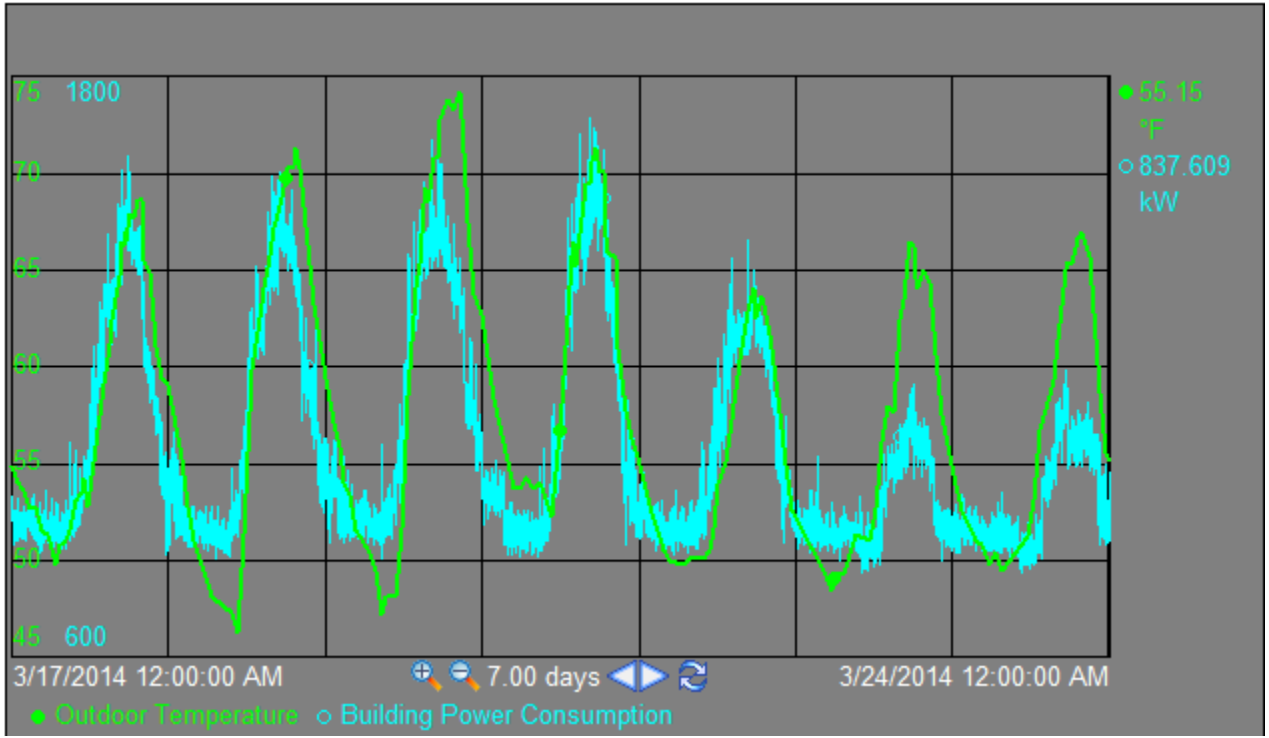


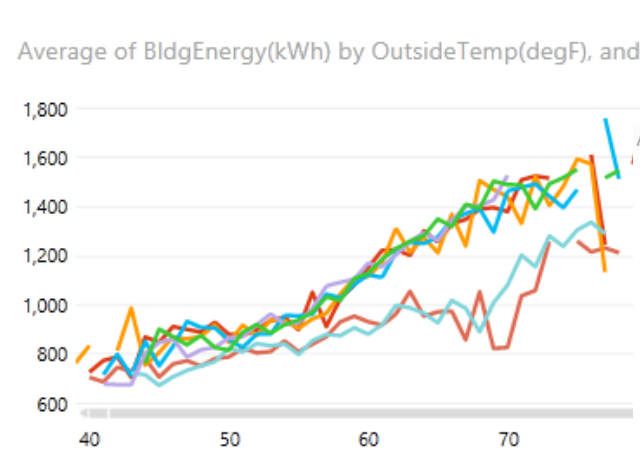
Business Intelligence / Visual Analytics

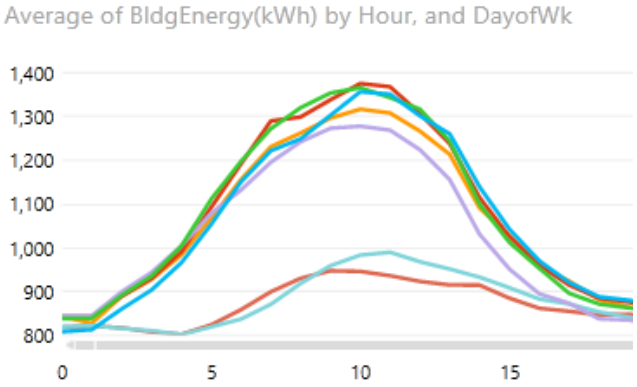


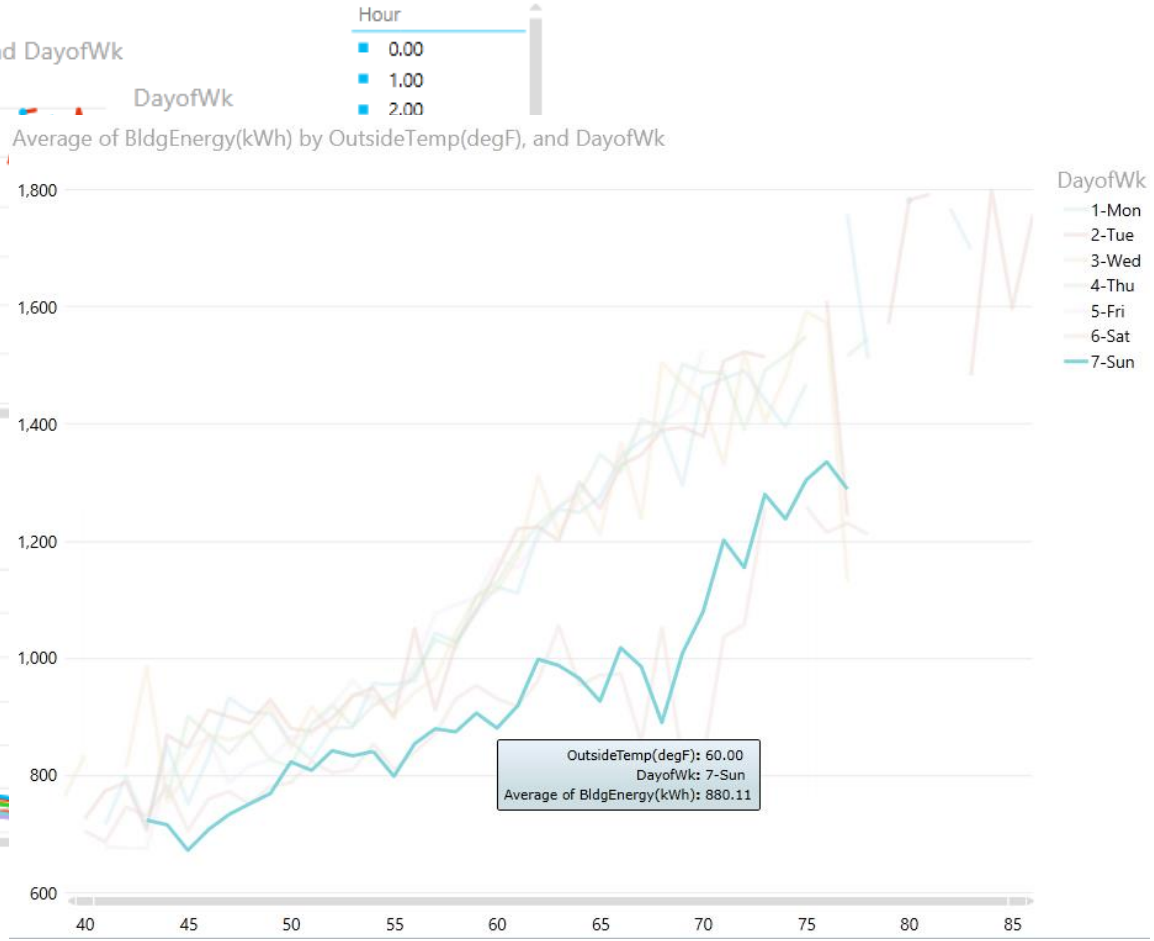
Machine Learning and Predictive Analytics

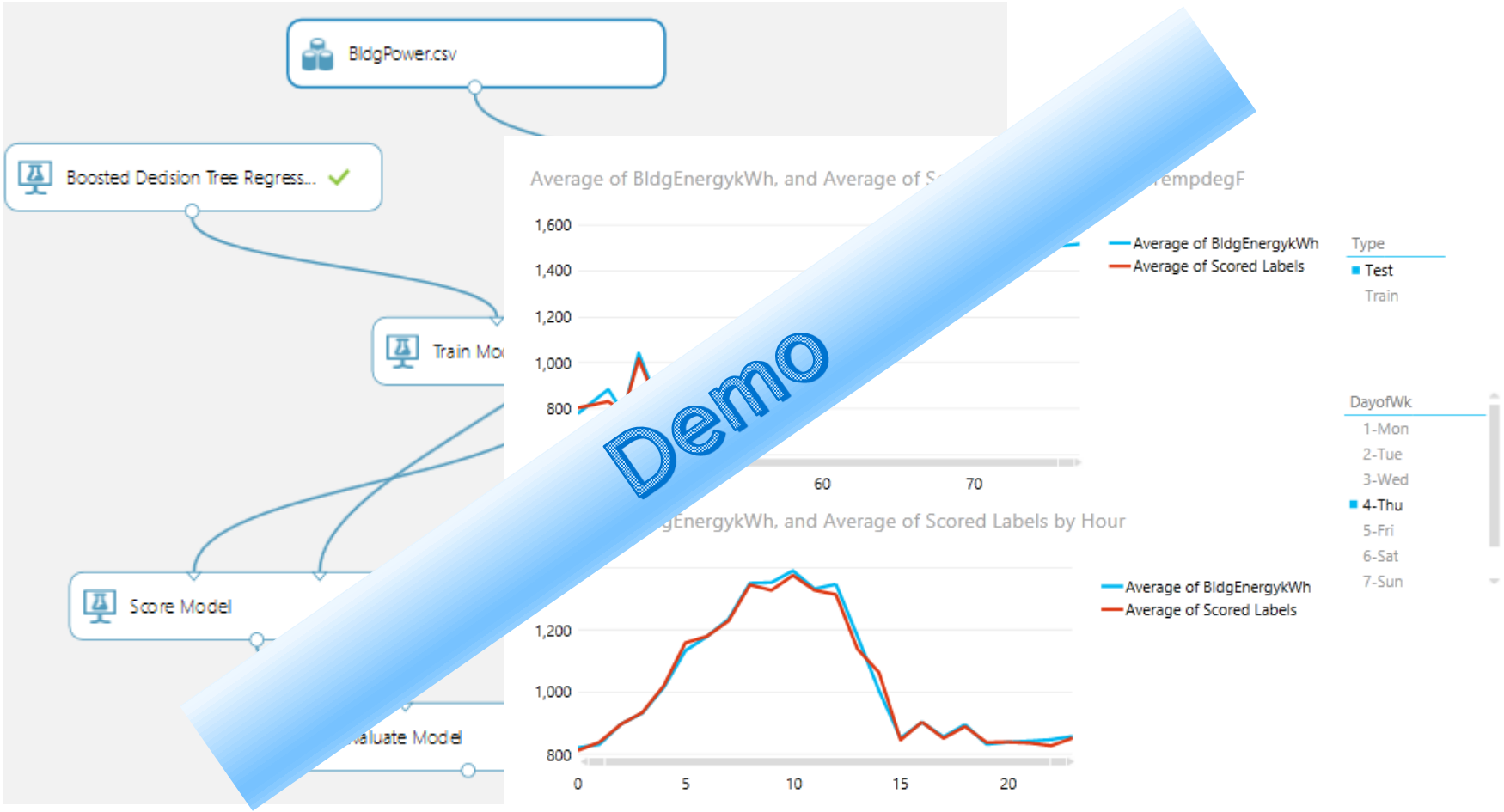


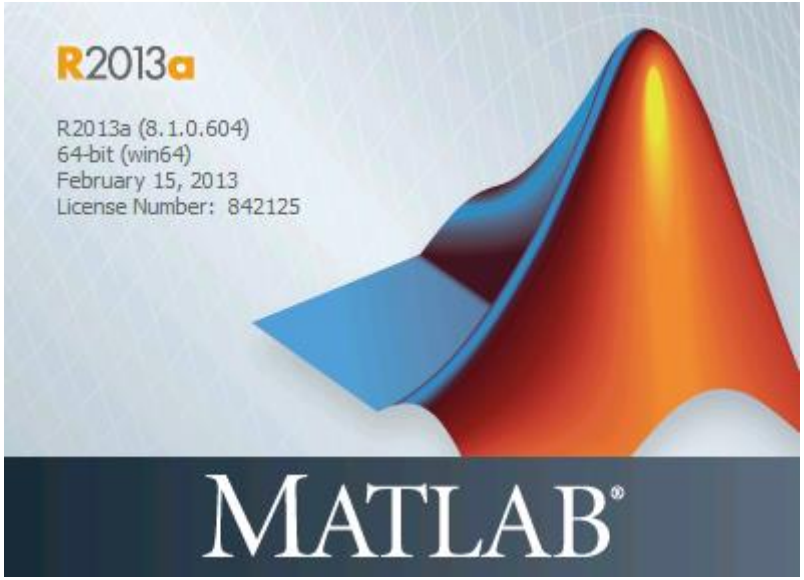




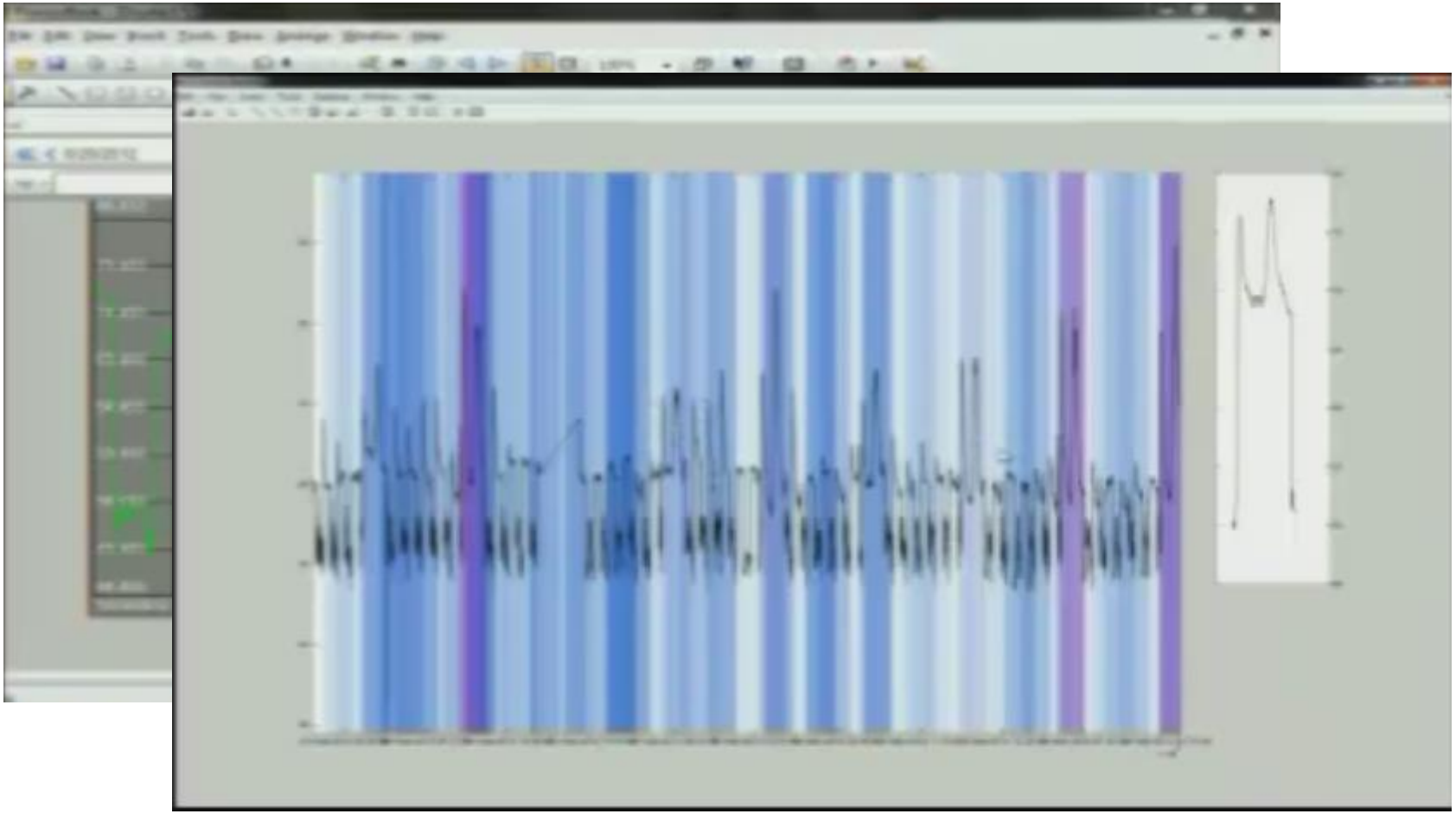




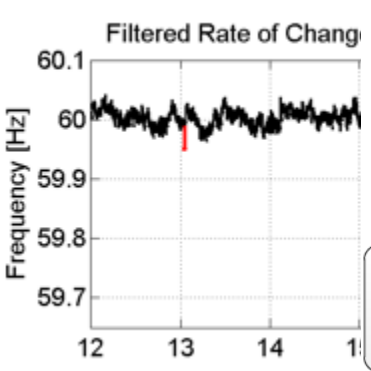


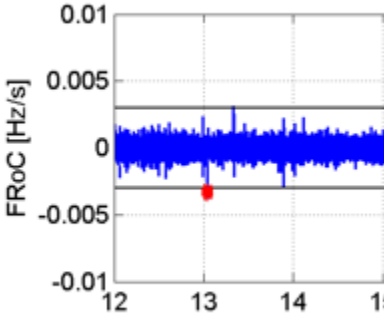


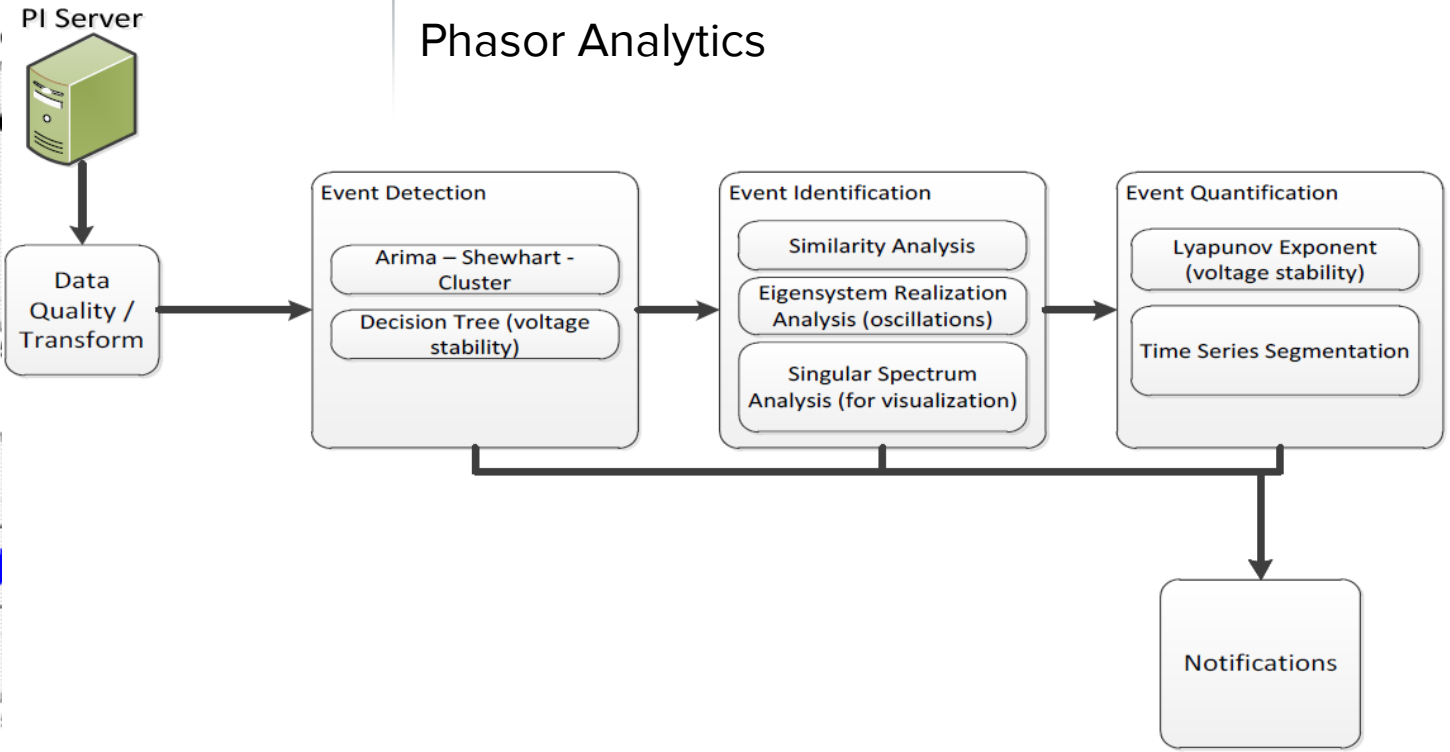
PI System with MATLAB...



Phasor Analytics







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

InformationWeek

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