“Rationalization of What and Where” for Advanced Analytics in an IIOT and Big Data World

Curt Hertler, Global Solutions Architect, OSIsoft

Teradata - *Data Warehouse*

- 1976–1979: concept of Teradata grows from research at ***California Institute of Technology*** (Caltech) and from the discussions of ***Citibank's*** advanced technology group.
- 1984: Teradata releases the world's ***first parallel data warehouses and data marts***.
- 1986: Fortune Magazine names Teradata ***"Product of the Year."***
- 1992: Teradata creates the first system over ***1 terabyte***, which goes live at ***Wal-Mart***.
- 1997: Teradata customer creates world's largest production database at ***24 terabytes***.
- 1999: Teradata customer has world's largest database with ***130 terabytes***.
- 2014: Teradata acquires Rainstor, a company specializing in online ***big data archiving on Hadoop***.

https://en.wikipedia.org/wiki/Teradata#History

Hadoop - *Wide Search of Everything*

- The genesis of Hadoop came from the ***Google*** File System paper that was published in ***October 2003***.

- This paper spawned another research paper from Google – MapReduce: Simplified Data Processing on Large Clusters.

- Development started in the Apache Nutch project, but was moved to the new Hadoop subproject in January 2006. Doug Cutting, who was working at ***Yahoo!*** at the time, ***named it after his son's toy elephant***.

https://en.wikipedia.org/wiki/Apache_Hadoop#History

In-Memory Columnar processing - *Tabular based analytics*

Company	Product	Description
IBM	Informix	Supports Dynamic In-memory (***in-memory columnar processing***) Parallel Vector Processing, Actionable Compression, and Data Skipping technologies, collectively called "Blink Technology" by IBM. Released: March 2011.
IBM	DB2 BLU	IBM DB2 for Linux, UNIX and Windows supports dynamic in-memory (***in-memory columnar processing***) parallel vector processing, actionable compression, and data-skipping technologies, collectively called IBM BLU Acceleration by IBM.
Microsoft	SQL Server	SQL Server 2012 included an ***in-memory*** technology called xVelocity ***column-store*** indexes targeted for data-warehouse workloads.
SAP	HANA	Short for 'High Performance Analytic Appliance' is an ***in-memory, column-oriented***, relational database management system written in C, C++.

https://en.wikipedia.org/wiki/List_of_in-memory_databases

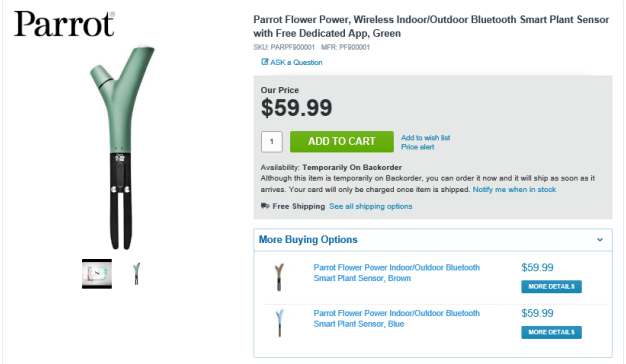
Power Pivot for Excel - *Self Service, Large Data Analytics*

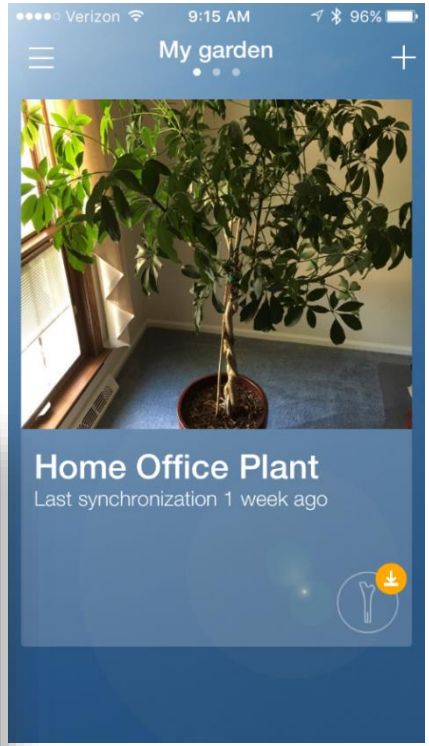
Power Pivot is a feature of Microsoft Excel. It is ***available as an add-in in Excel 2010 and 2013, and is included natively in Excel 2016***. PowerPivot extends a local instance of Microsoft Analysis Services Tabular that is embedded directly into an Excel Workbook.

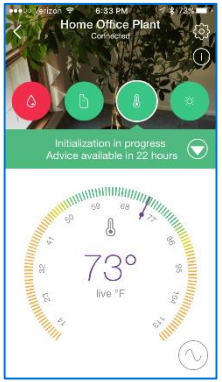
PowerPivot uses the SSAS Vertipaq compression engine to hold the data model ***in memory on the client computer***. Practically, this means that PowerPivot is acting as an ***Analysis Services Server instance on the local workstation***.

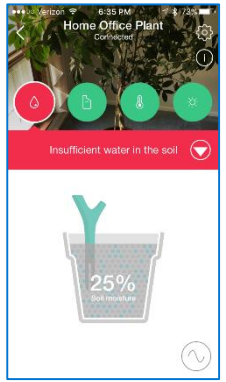
https://en.wikipedia.org/wiki/Power_Pivot

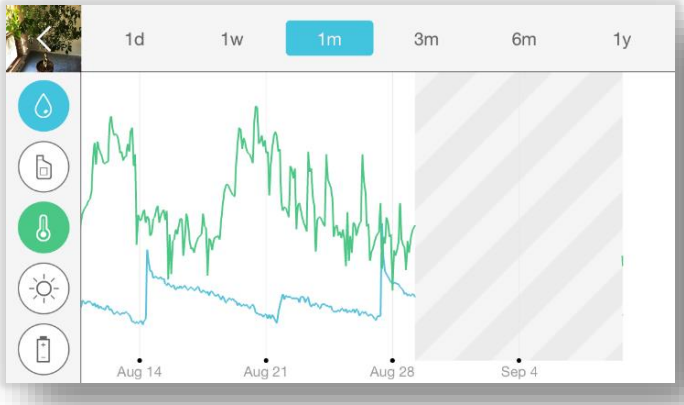
Internet of Things - *This means all things!*

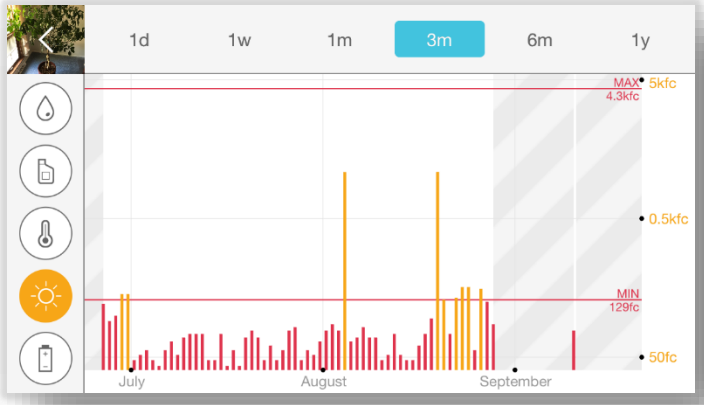






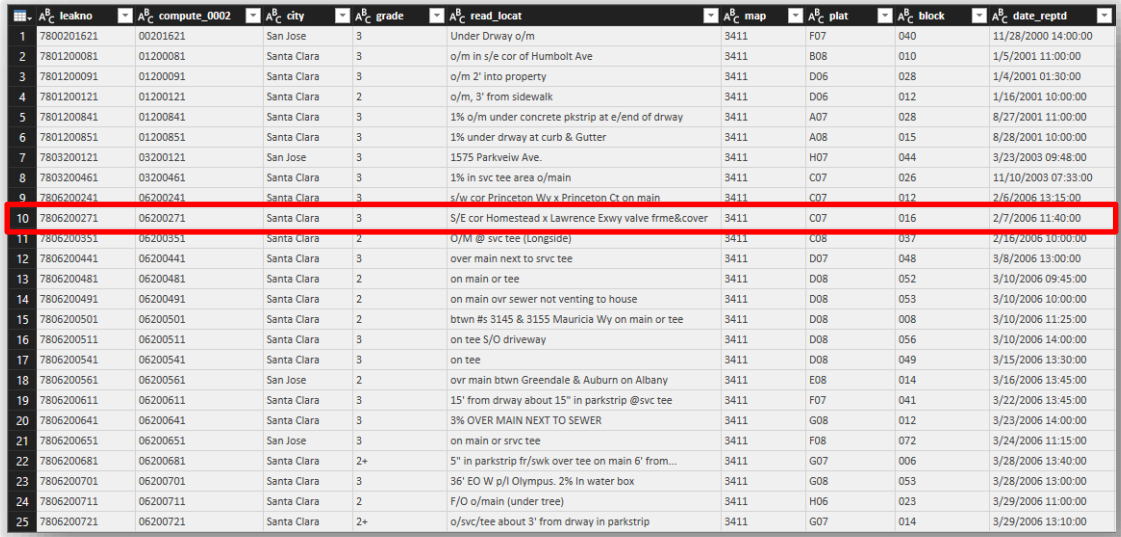






Real Time Data is Different

- Transactional data is recorded in a tabular format with values associated by columns in each row.



	leakno	compute_0002	city	grade	read_locat	map	plat	block	date_reptd
1	7800201621	00201621	San Jose	3	Under Drway o/m	3411	F07	040	11/28/2000 14:00:00
2	7801200081	01200081	Santa Clara	3	o/m in s/e cor of Humbolt Ave	3411	B08	010	1/5/2001 11:00:00
3	7801200091	01200091	Santa Clara	3	o/m 2' into property	3411	D06	028	1/4/2001 01:30:00
4	7801200121	01200121	Santa Clara	2	o/m, 3' from sidewalk	3411	D06	012	1/16/2001 10:00:00
5	7801200841	01200841	Santa Clara	3	1% o/m under concrete pkstrip at e/end of drway	3411	A07	028	8/27/2001 11:00:00
6	7801200851	01200851	Santa Clara	3	1% under drway at curb & Gutter	3411	A08	015	8/28/2001 10:00:00
7	7803200121	03200121	San Jose	3	1575 Parkveiw Ave.	3411	H07	044	3/23/2003 09:48:00
8	7803200461	03200461	Santa Clara	3	1% in svc tee area o/main	3411	C07	026	11/10/2003 07:33:00
9	7806200241	06200241	Santa Clara	3	s/w cor Princeton Wy x Princeton Ct on main	3411	C07	012	2/6/2006 13:15:00
10	7806200271	06200271	Santa Clara	3	S/E cor Homestead x Lawrence Exwy valve frme&cover	3411	C07	016	2/7/2006 11:40:00
11	7806200351	06200351	Santa Clara	2	O/M @ svc tee (Longside)	3411	C08	037	2/16/2006 10:00:00
12	7806200441	06200441	Santa Clara	3	over main next to srvc tee	3411	D07	048	3/8/2006 13:00:00
13	7806200481	06200481	Santa Clara	2	on main or tee	3411	D08	052	3/10/2006 09:45:00
14	7806200491	06200491	Santa Clara	2	on main ovr sewer not venting to house	3411	D08	053	3/10/2006 10:00:00
15	7806200501	06200501	Santa Clara	2	btwn #s 3145 & 3155 Mauricia Wy on main or tee	3411	D08	008	3/10/2006 11:25:00
16	7806200511	06200511	Santa Clara	3	on tee S/O driveway	3411	D08	056	3/10/2006 14:00:00
17	7806200541	06200541	Santa Clara	3	on tee	3411	D08	049	3/15/2006 13:30:00
18	7806200561	06200561	San Jose	2	ovr main btwn Greendale & Auburn on Albany	3411	E08	014	3/16/2006 13:45:00
19	7806200611	06200611	Santa Clara	3	15' from drway about 15" in parkstrip @svc tee	3411	F07	041	3/22/2006 13:45:00
20	7806200641	06200641	Santa Clara	3	3% OVER MAIN NEXT TO SEWER	3411	G08	012	3/23/2006 14:00:00
21	7806200651	06200651	San Jose	3	on main or srvc tee	3411	F08	072	3/24/2006 11:15:00
22	7806200681	06200681	Santa Clara	2+	5" in parkstrip fr/swk over tee on main 6' from...	3411	G07	006	3/28/2006 13:40:00
23	7806200701	06200701	Santa Clara	3	36' EO W p/l Olympus. 2% In water box	3411	G08	053	3/28/2006 13:00:00
24	7806200711	06200711	Santa Clara	2	F/O o/main (under tree)	3411	H06	023	3/29/2006 11:00:00
25	7806200721	06200721	Santa Clara	2+	o/svc/tee about 3' from drway in parkstrip	3411	G07	014	3/29/2006 13:10:00

- Real-time data is recorded with only time context, i.e. value and timestamp.

56.902 03-SEP-2016 11:23 AM

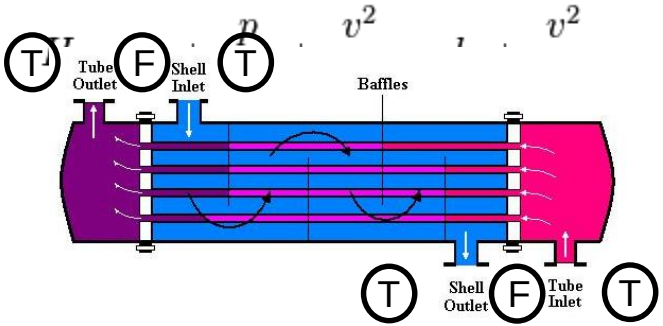
Real-time Data Requires Context

Time

63.781 03-SEP-2016 11:19 AM → 56.902 03-SEP-2016 11:23 AM → 58.341 03-SEP-2016 11:41 AM

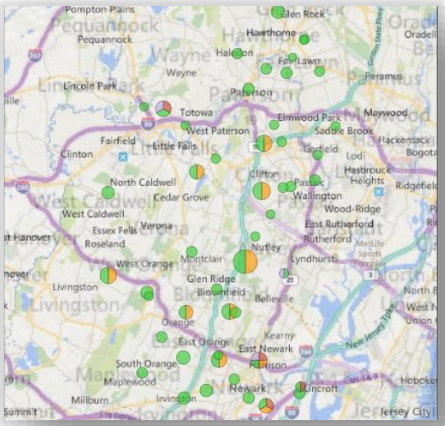
Asset

Scientific Fact



Daniel Bernoulli
(1700 – 1782)

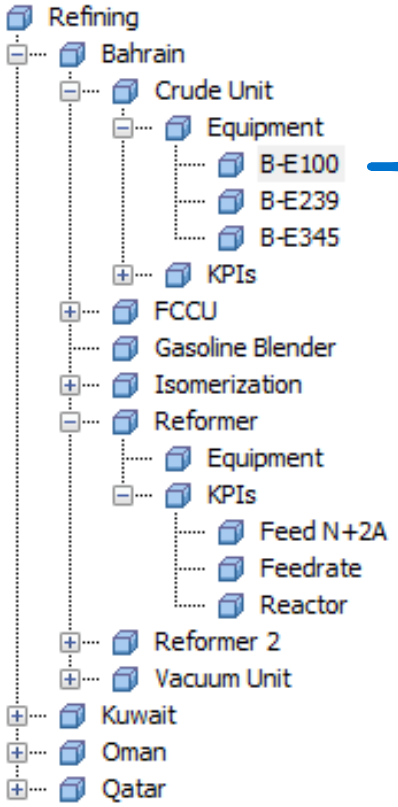
Location



PI AF Supplies Required Context

Asset Hierarchy

- Plant
- Process
- Assets
- Process Context
- Location
- Specifications



Process Context

Category: Process Data	
Cold Side Inlet Temperature	77.1157989501953 °F
Cold Side Outlet Temperature	131.192291259766 °F
Hot Side Inlet Temperature	374.601501464844 °F
Hot Side Outlet Temperature	292.926361083984 °F

Location

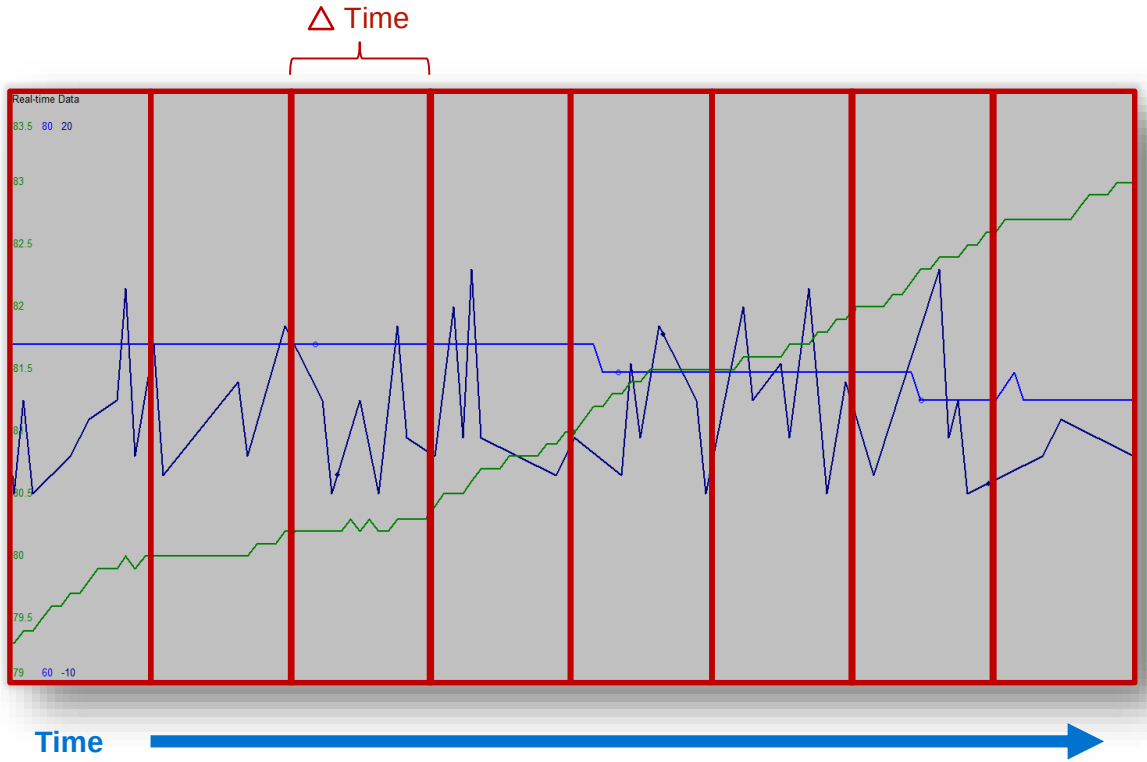
Category: Location	
Address	2265 W Salinas St, San ...
Latitude	29.43027
Longitude	-98.518172

Specifications

Category: Specifications	
Area	1200 ft2
Coefficient	75.66 BTU per F ft2 Hr
Service	Crude vs. Naphtha

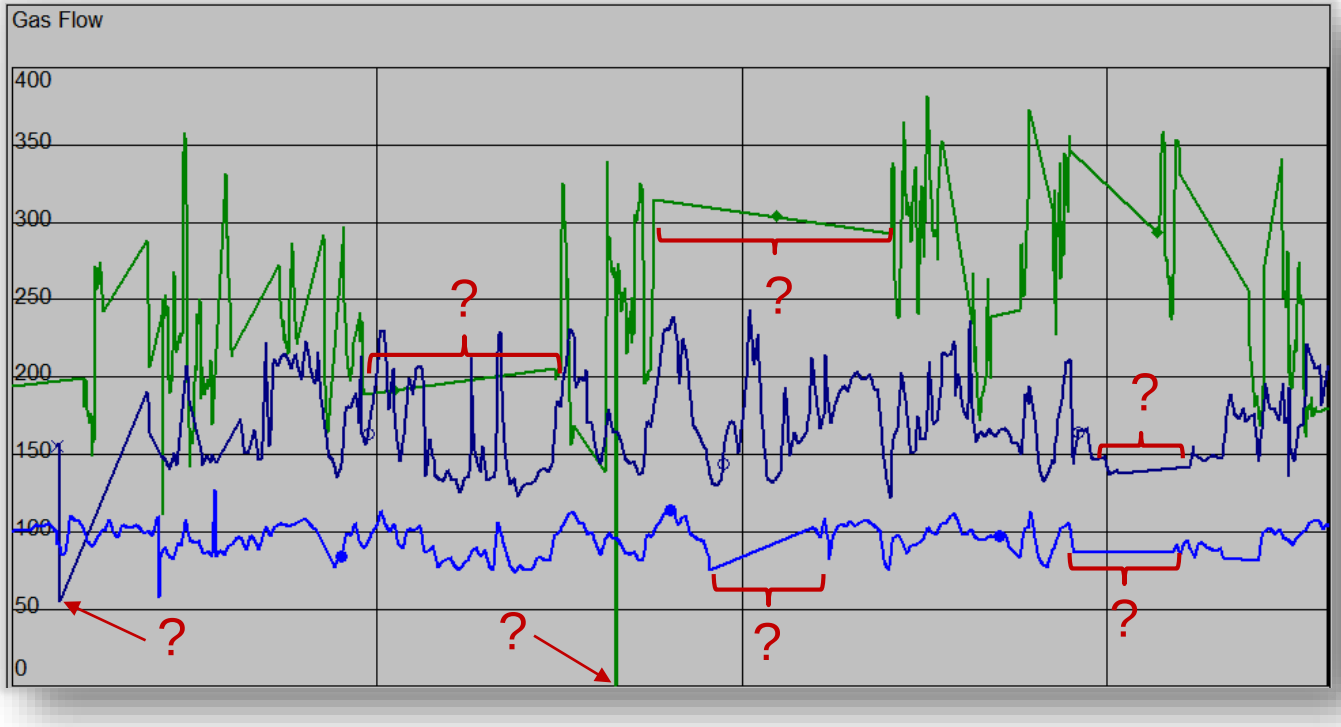
Real-time Data Requires Proper Aggregation

- Optimized for Operations, along the time-dimension for agile performance and access.
- Interpolations and time-weighted aggregations are required to fit tabular formatting requirements.



Real-time Data Never Stops

- Equipment failure ?
- Sensor failure ?
- Process upset ?

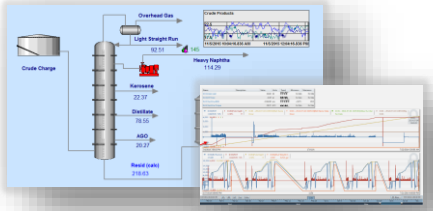


Enabling Analytics for Operational Intelligence

Real-Time Decision Analysis

Retrospective & Predictive Analysis

PI Integrator for Business Analytics

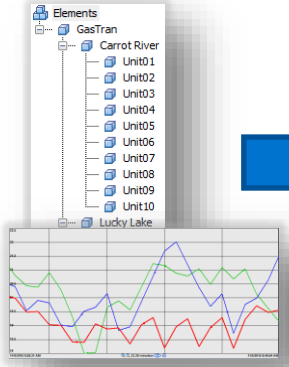


Time and Event
Trending & Awareness

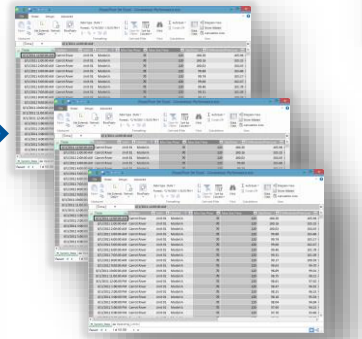
Specialized Models
Simulation & Optimization

$$Q = \frac{\Delta P_{DD} * kh}{141.2 \mu B_0 \left\{ ln \frac{r_e}{r_w} - \frac{3}{4} + S \right\}}$$

Descriptive
Condition & Performance

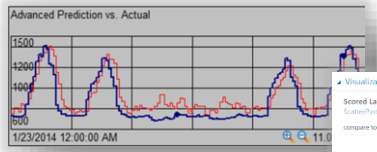


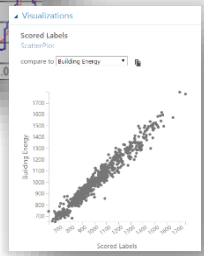
Time, Event and Asset Context



Tabular Context

Visual
Dashboards &
Multidimensional Assessment





Predictive
Statistical Modelling
& Machine Learning

Common Ground between Technological Contexts

Descriptive Analytics – Condition & *Performance*

- First Principle Relationships that ***always*** exists between process measurements.
- Enables real-time decision making only when visible, i.e. not performed in spreadsheets.
- Operations ownership requires transparency of methods, assumptions, and frequency.

Daniel Bernoulli
(1700 – 1782)

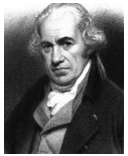
$$H = z + \frac{p}{\rho g} + \frac{v^2}{2g} = h + \frac{v^2}{2g}$$

Benoît Clapeyron
(1799 – 1864)

$$Q = \frac{\Delta P_{DD} * kh}{141.2 \mu B_0 \left\{ ln \frac{r_e}{r_w} - \frac{3}{4} + S \right\}}$$

$$PV = nRT$$

$$F = \frac{(R+1)^{1/2} \text{ x ln } ((1-SR)/(1-S))}{(1-R) \text{ x ln } \left\{ \frac{2-S(R+1-(R+1)^{1/2})}{2-S(R+1+(R+1)^{1/2})} \right\}}$$

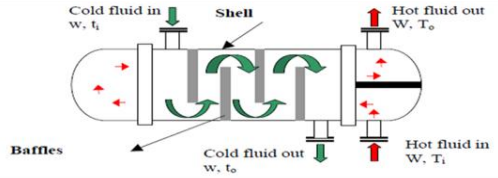


James Watt
(1736–1819)

$$P = IV = I^2R = \frac{V^2}{R}$$

First Principles Analytics - *PI Asset Analytics*

- *Configure* calculations at scale
- Math, statistical, logical and steam table functions
- Supports basic predictive analytics
- Supports future data for forecasting
- **Backfill ! Backfill ! Backfill !**



Heat Exchanger Key Performance Indicator:

Overall heat transfer coefficient

$$U = \frac{Q}{A \times \text{Corrected LMTD}}$$

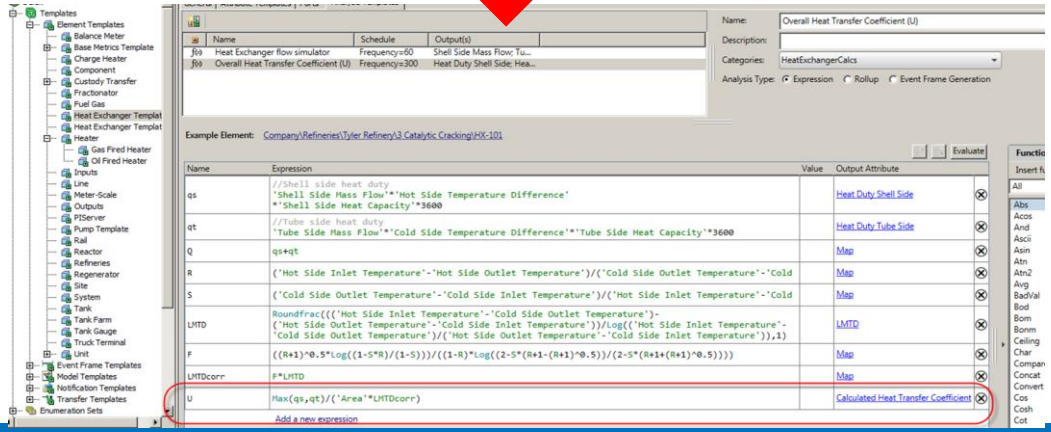
RULE: *IF the heat transfer coefficient is decreasing, THEN the* ***Heat Exchanger FOULING*** *!!! Cleaining is required!*

Calculation Steps:

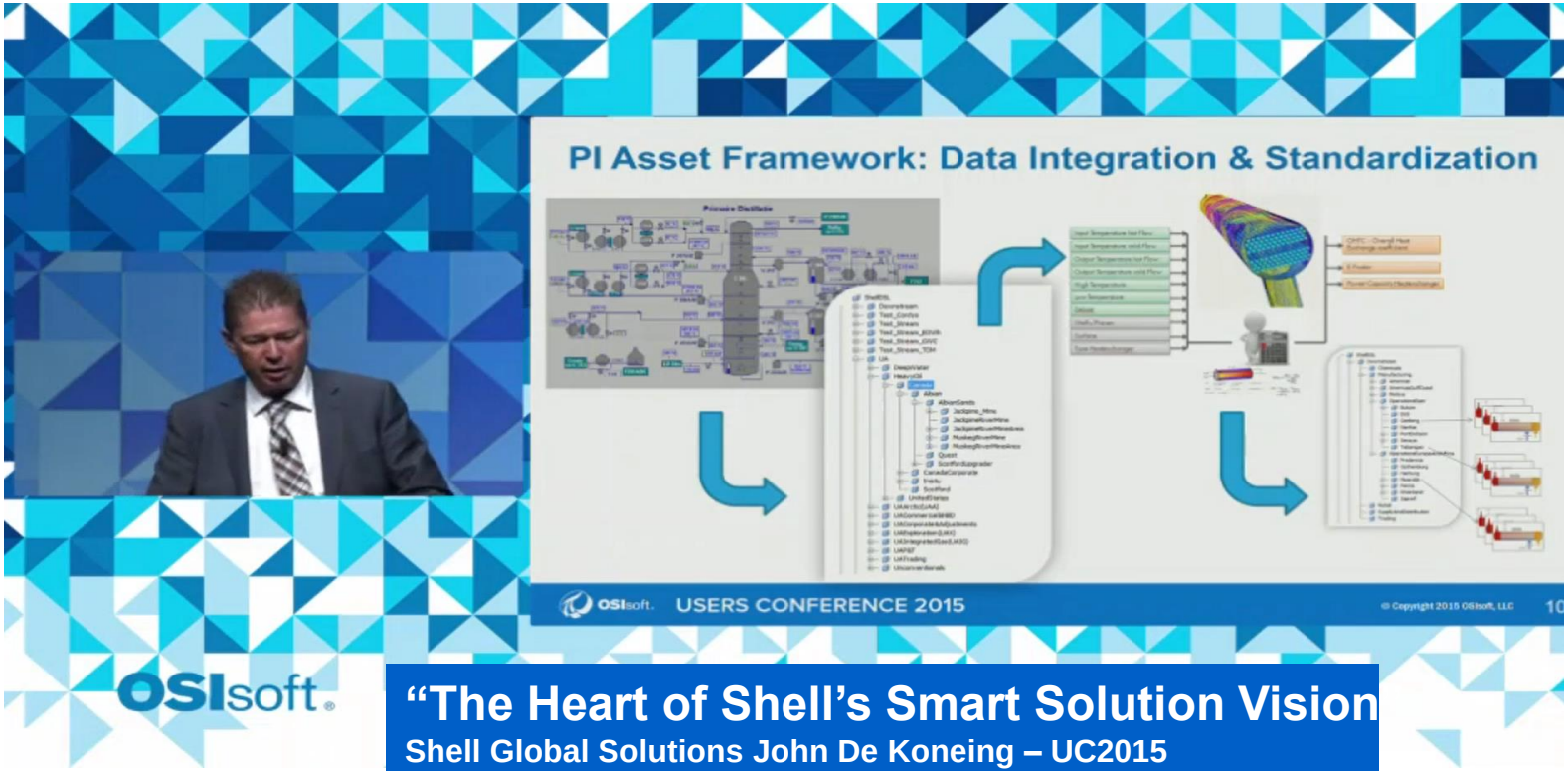
1. Heat Duty, $Q = q_s + q_t$
   $q_s = W \times C_{ph} \times (T_i - T_o)/1000/3600$
   $q_s = w \times C_{pc} \times (t_o - t_i)/1000/3600$
2. Hot Fluid Pressure Drop, $\Delta P_h = P_i - P_o$
3. Cold fluid pressure drop, $\Delta P_c = p_i - p_o$
4. Temperature range hot fluid, $\Delta T = T_i - T_o$
5. Temperature range cold fluid, $\Delta t = t_o - t_i$
6. Capacity ratio, $R = W \times C_{Ph} / w \times C_{pc}$ (or) $(T_i - T_o) / (t_o - t_i)$
7. Effectiveness, $S = (t_o - t_i) / (T_i - t_i)$
8. LMTD
   LMTD Counter current Flow = $((T_i - t_o) - (T_o - t_i)) / \ln ((T_i - t_o)/(T_o - t_i))$
   LMTD Co current Flow = $((T_i - t_i) - (T_o - t_o)) / \ln ((T_i - t_i)/(T_o - t_o))$
   Correction factor for LMTD to account for Cross flow

$$F = \frac{(R+1)^{1/2} \times \ln((1-SR)/(1-S))}{(1-R) \times \ln\left\{\frac{2-S(R+1-(R+1)^{1/2})}{2-S(R+1+(R+1)^{1/2})}\right\}}$$

9. Corrected LMTD = F x LMTD

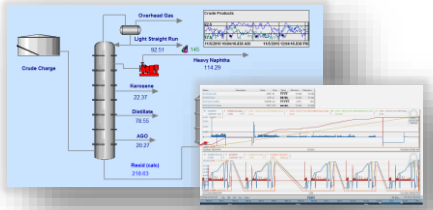


500 PI AF Templates “Smart Asset” Templates



Enabling Analytics for Operational Intelligence

Real-Time Decision Analysis



Time and Event
Trending & Awareness

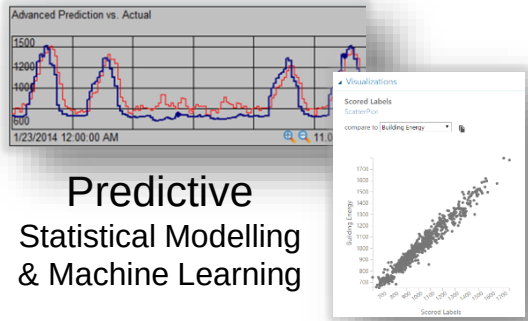
Specialized Models
Simulation & Optimization

$$Q = \frac{\Delta P_{DD} * kh}{141.2 \mu B_0 \left\{ ln \frac{r_e}{r_w} - \frac{3}{4} + S \right\}}$$

Descriptive
Condition & Performance

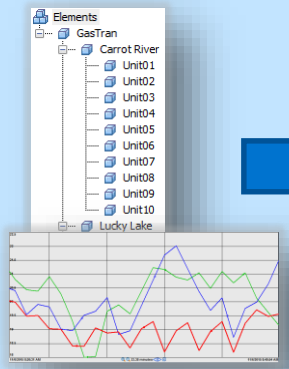
Retrospective & Predictive Analysis

Visual
Dashboards &
Multidimensional Assessment

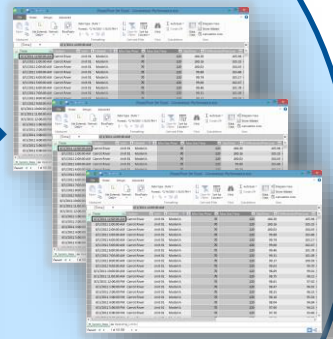


Predictive
Statistical Modelling
& Machine Learning

PI Integrator for Business Analytics



Time, Event and Asset Context

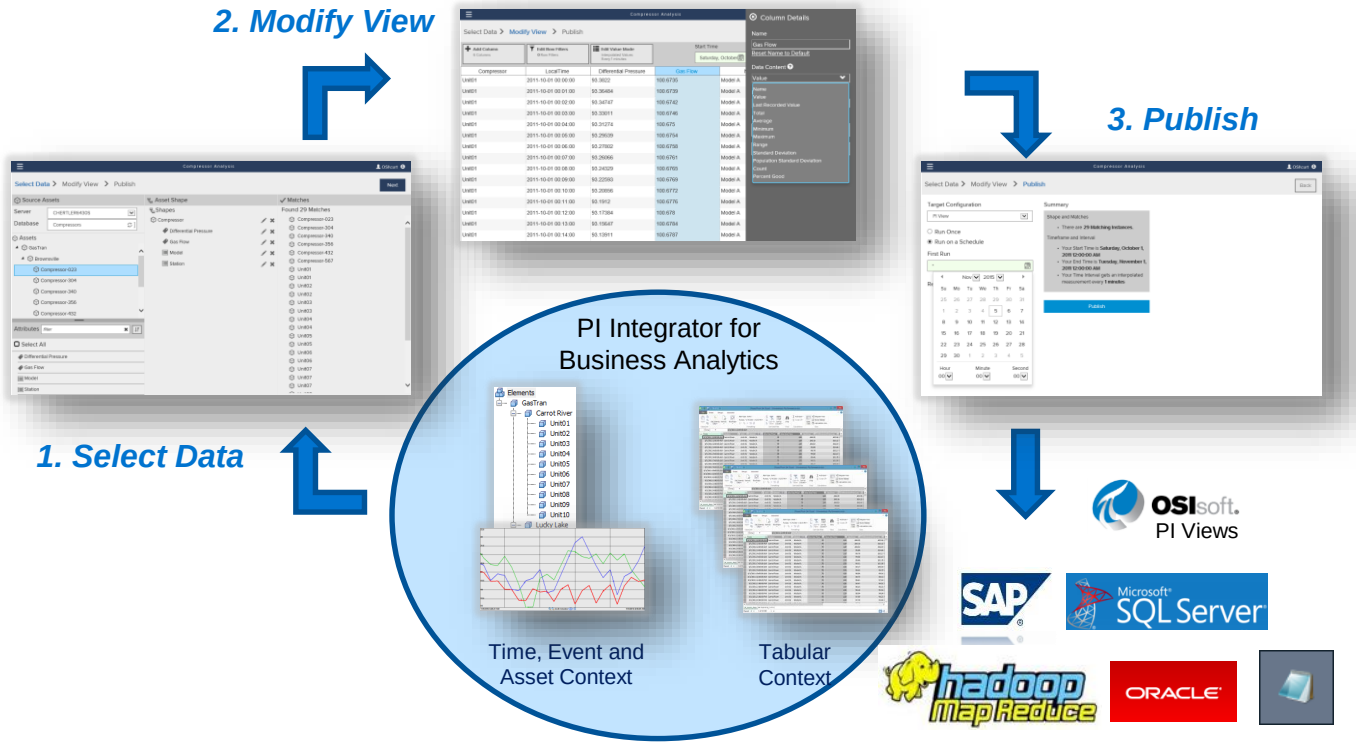


Tabular Context

PI Integrator for Business Analytics

Easy, scalable way for users to create contextualized views of operational data.

- **Select** assets and their attributes from an AF hierarchy.
- **Modify View** by setting time range, row interval, and column aggregations. Add filtering rules to “cleanse” data.
- **Publish** once or on a scheduled bases.

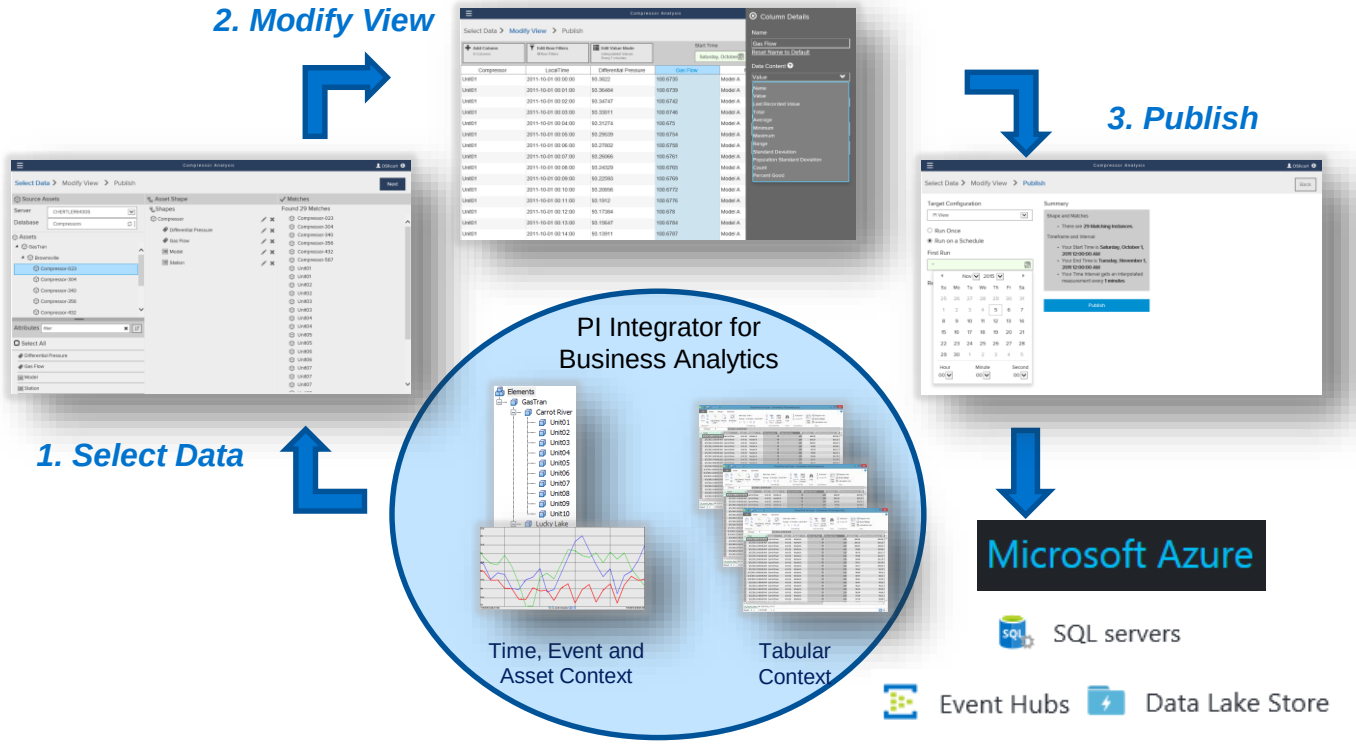


PI Integrator for Microsoft Azure – *Coming Soon!*

Easy, scalable way for users to create contextualized views of operational data.

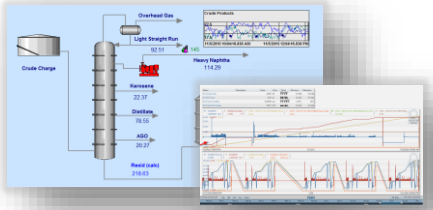
- **Select** assets and their attributes from an AF hierarchy.
- **Modify View** by setting time range, row interval, and column aggregations.

  Add filtering rules to "cleanse" data.
- **Publish** once or on a scheduled bases.



Enabling Analytics for Operational Intelligence

Real-Time Decision Analysis



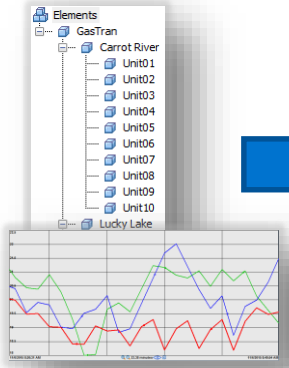
Time and Event
Trending & Awareness

Specialized Models
Simulation & Optimization

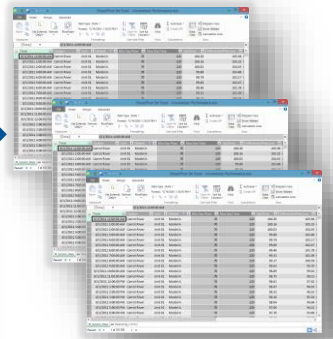
$$Q = \frac{\Delta P_{DD} * kh}{141.2 \mu B_0 \left\{ ln \frac{r_e}{r_w} - \frac{3}{4} + S \right\}}$$

Descriptive
Condition & Performance

PI Integrator for Business Analytics



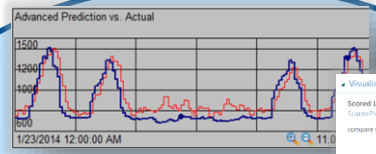
Time, Event and Asset Context

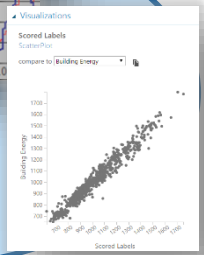


Tabular Context

Retrospective & Predictive Analysis

Visual
Dashboards & Multidimensional Assessment



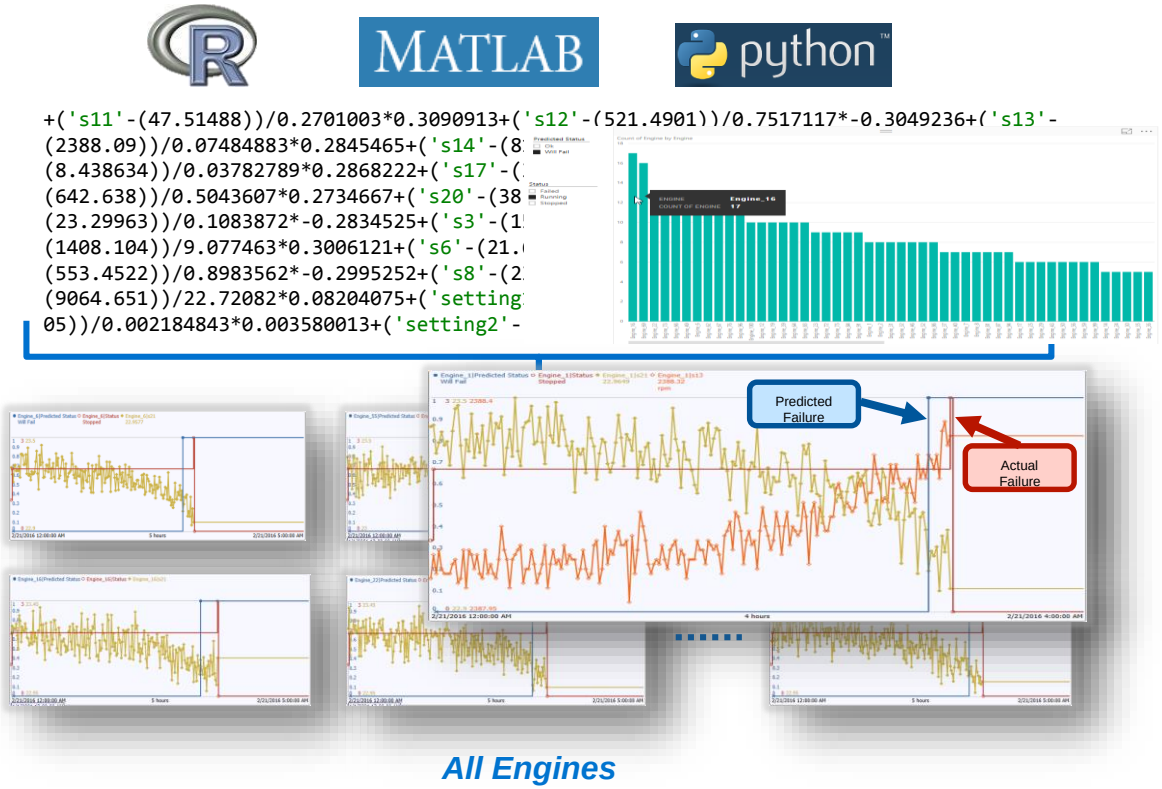


Predictive
Statistical Modelling & Machine Learning

Statistical Modelling - *Predict Asset Failure*

Complex systems descriptive equations are too numerous and interrelated.

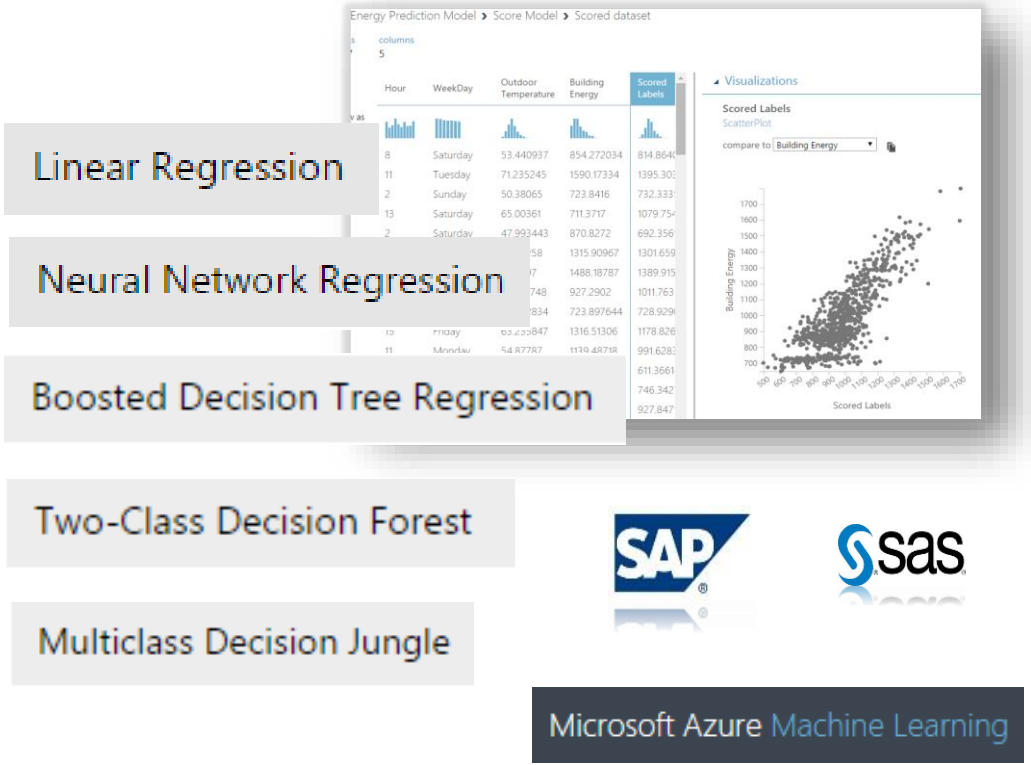
- Create an operationalized model to reduce unplanned downtime for 100 engines.
- ***PI Integrator for BA*** used to extract data for 2,300 sensors leading up to engine failures.
- Developed a statistical model using R for predicting failure.
- Tested and operationalized using PI Analytics for all engines.



Predictive Analytics - *Machine Learning*

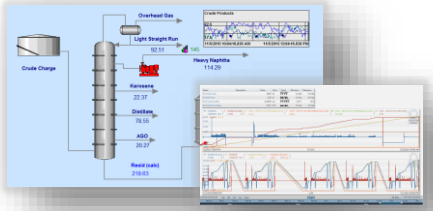
Machine learning improves statistical model by “learning” from additional operating data.

- OSIsoft Partners provide statistical applications and data science services.
- Gain business insights from datasets coming from many sources, e.g. data warehouse.
- **Operationalization supported by scheduled publication from PI Integrator for Business Analytics.**



Enabling Analytics for Operational Intelligence

Real-Time Decision Analysis



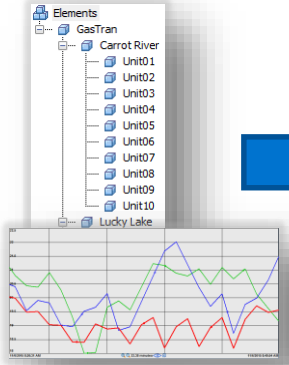
Time and Event
Trending & Awareness

Specialized Models
Simulation & Optimization

$$Q = \frac{\Delta P_{DD} * kh}{141.2 \mu B_0 \left\{ ln \frac{r_e}{r_w} - \frac{3}{4} + S \right\}}$$

Descriptive
Condition & Performance

PI Integrator for Business Analytics

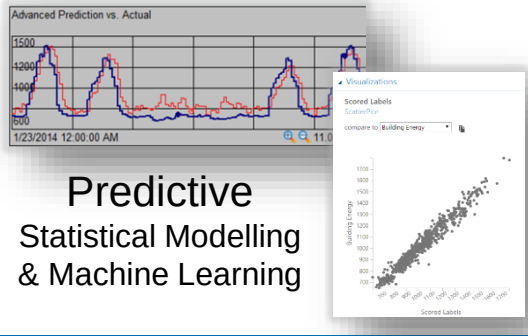


Time, Event and Asset Context

Tabular Context

Retrospective & Predictive Analysis

Visual
Dashboards & Multidimensional Assessment



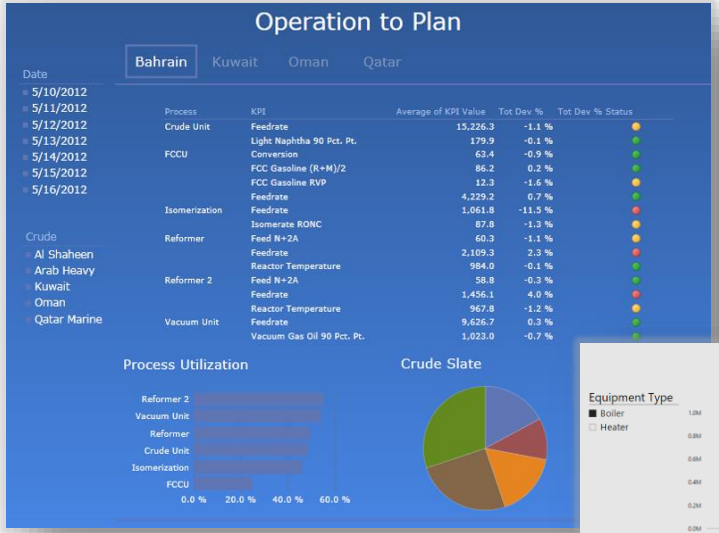
Predictive
Statistical Modelling & Machine Learning

Visual Analytics - Dashboarding & Reporting

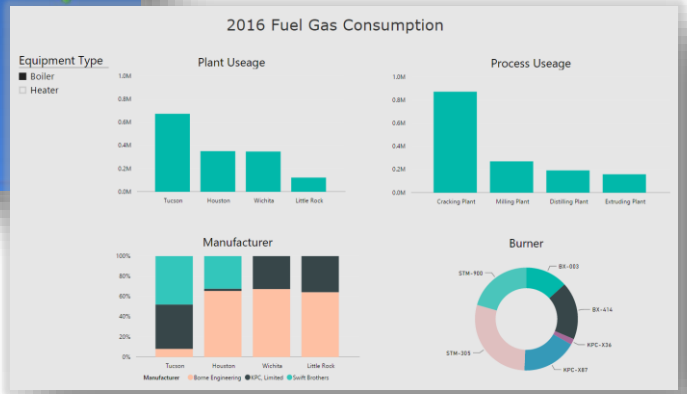
Dashboards and reports for performance assessment or accountability.

- Cross filtering charts for ad hoc investigation.
- **“What is shown in the report, stays in the report.”**
- Important aggregations can be permanently recorded in PI using AF Analytics.

Enterprise Scorecard



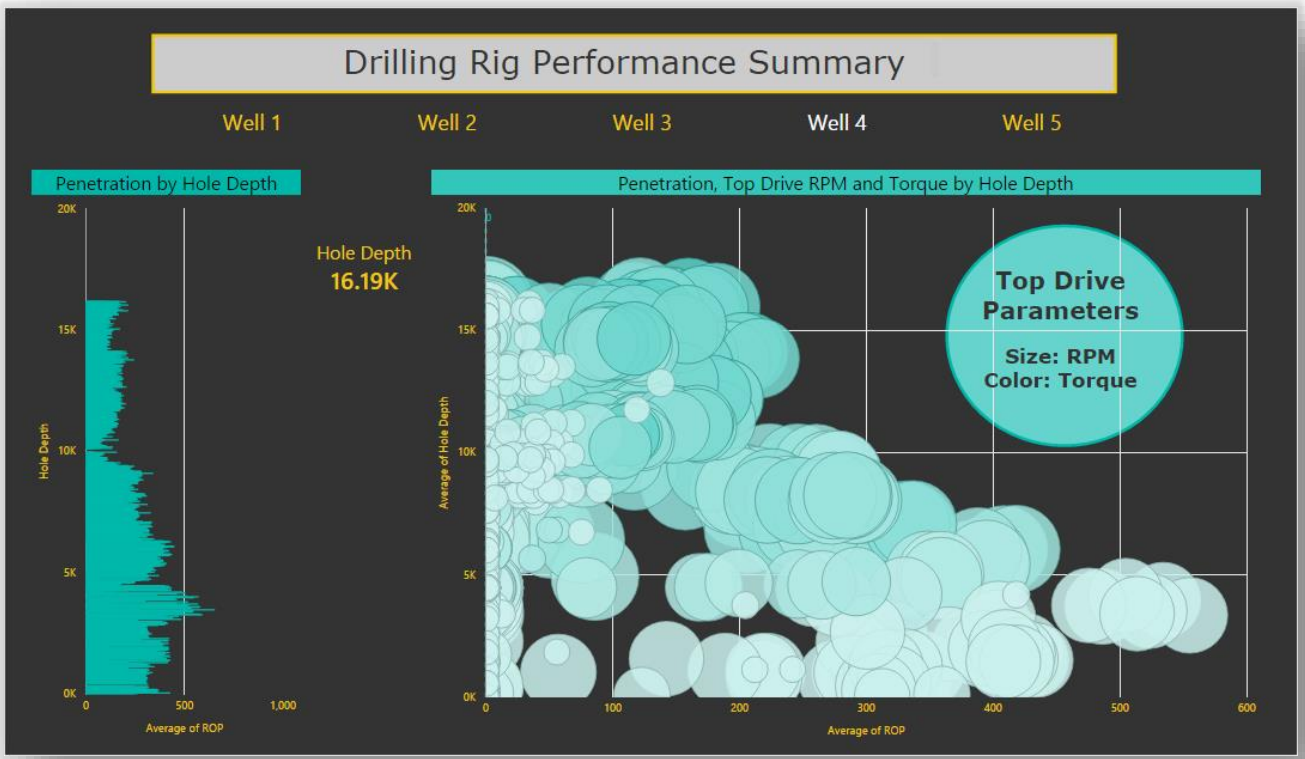
Energy Accounting



Visual Analytics - Multidimensional Assessment

Data collected to manage drilling operation is used to gain insights about formation geology.

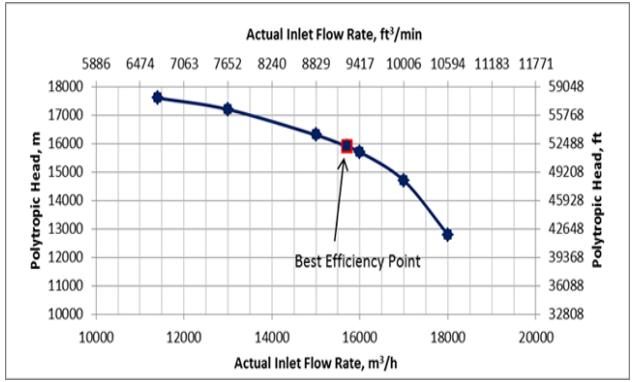
- Ad hoc analysis on any dimension, well depth, drilling rate, rpm and torque.
- Wide variety of charting objects available.

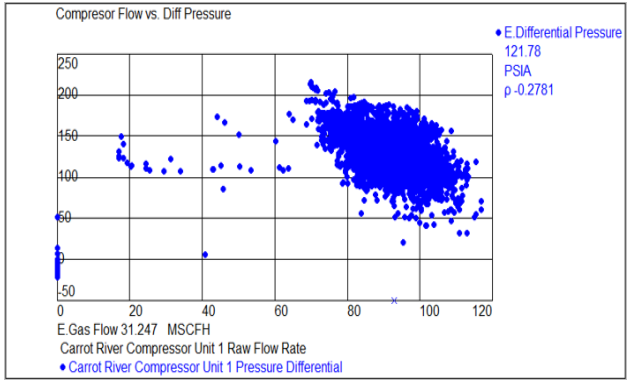


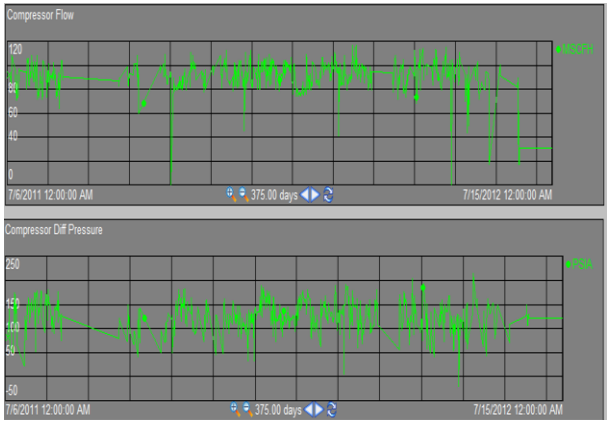
Visual Analytics – Asset Benchmarking

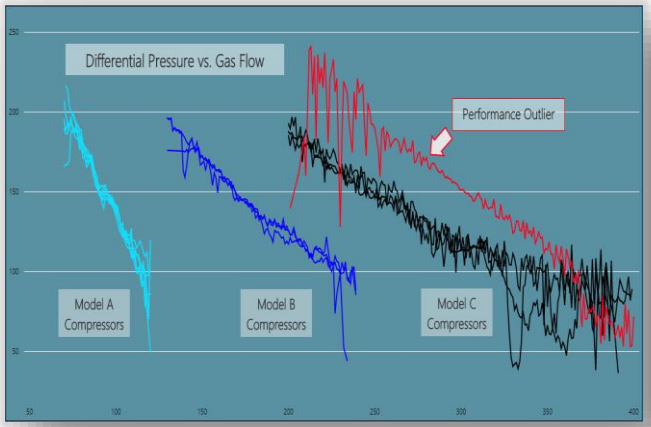
Benchmark similar assets against know performance characteristics.

- Real-time view is essential for current operation.
- Different tools required to analyze groups of assets.
- Summarize months of actual operations exposing actual performance profile.





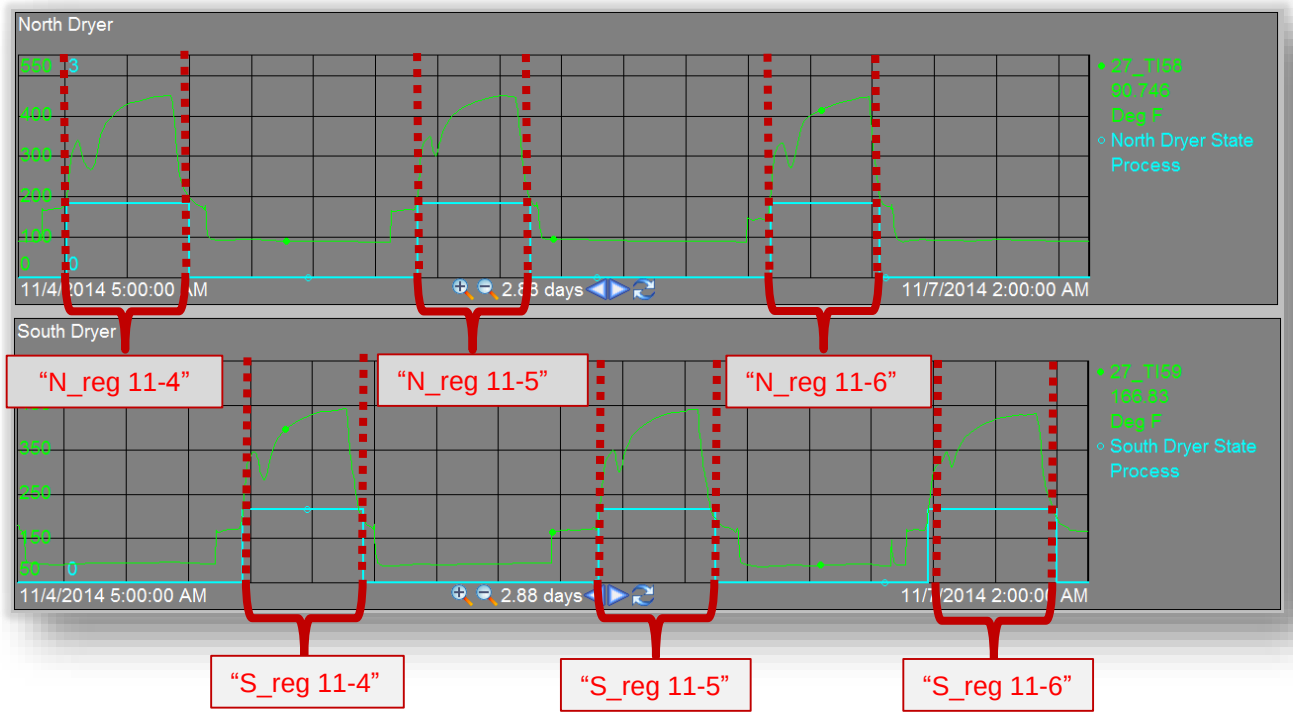




Event Frames – Feed Dryer Regeneration Cycles

PI Event Frames identify and record interesting periods of process operation for further analysis.

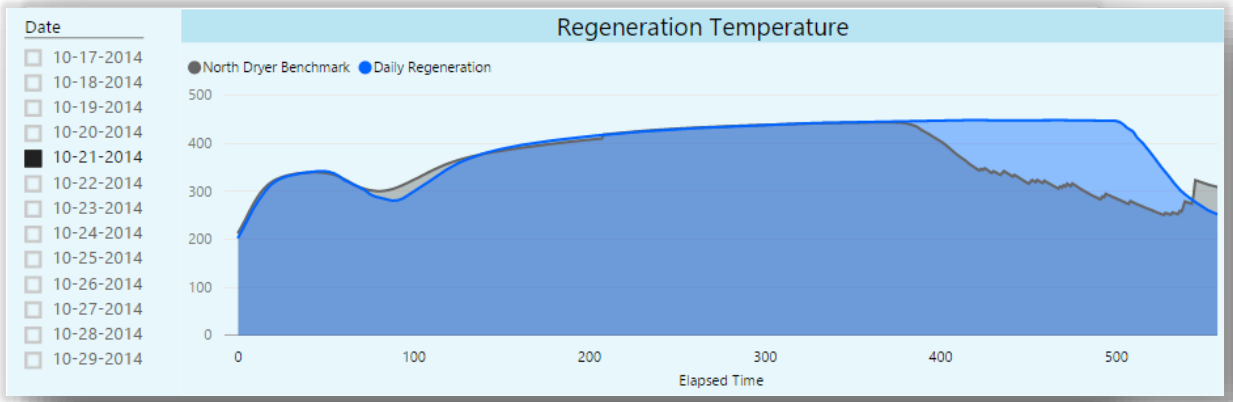
- Dryers cycle between processing and regeneration modes.
- PI Event Frames define regeneration cycles for each feed dryer.

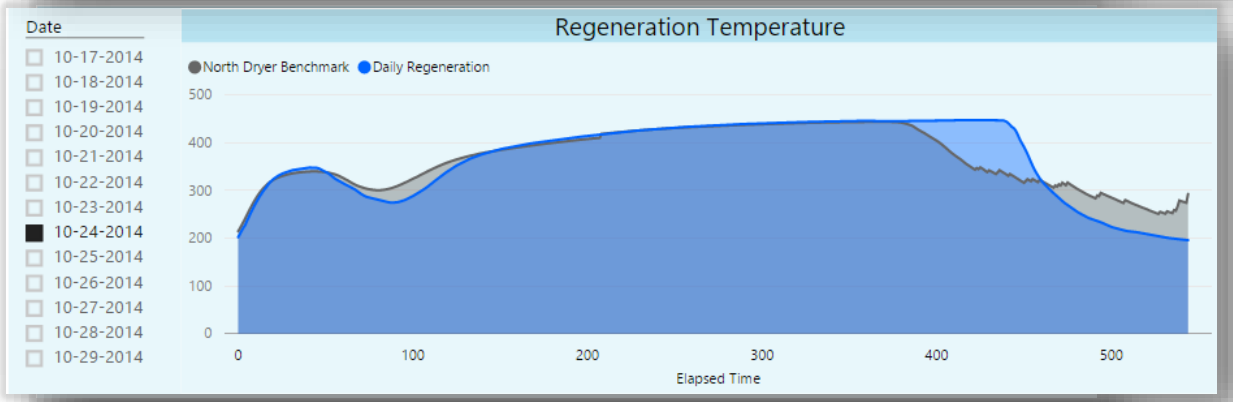


Visual Analytics - Conditional Profiling of Process Events

Feed dryer regeneration profile comparing two identical assets.

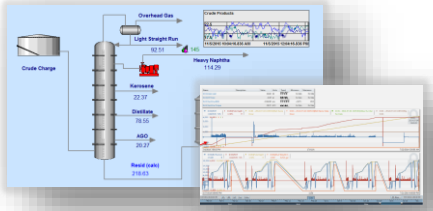
- PI Event Views can contain sampled process values throughout event.
- Profiled view of operation based on elapsed time.
- Conditional filtering of profile against benchmark or norm.





Enabling Analytics for Operational Intelligence

Real-Time Decision Analysis



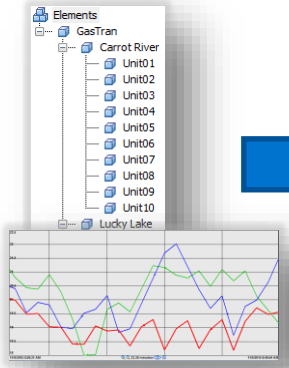
Time and Event
Trending & Awareness

Specialized Models
Simulation & Optimization

$$Q = \frac{\Delta P_{DD} * kh}{141.2 \mu B_0 \left\{ ln \frac{r_e}{r_w} - \frac{3}{4} + S \right\}}$$

Descriptive
Condition & Performance

PI Integrator for Business Analytics

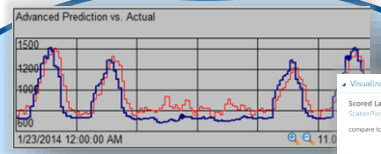


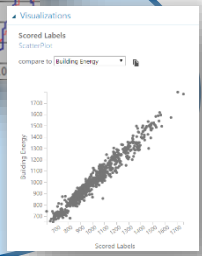
Time, Event and Asset Context

Tabular Context

Retrospective & Predictive Analysis

Visual
Dashboards & Multidimensional Assessment





Predictive
Statistical Modelling & Machine Learning

Integration with R Open Source Visuals

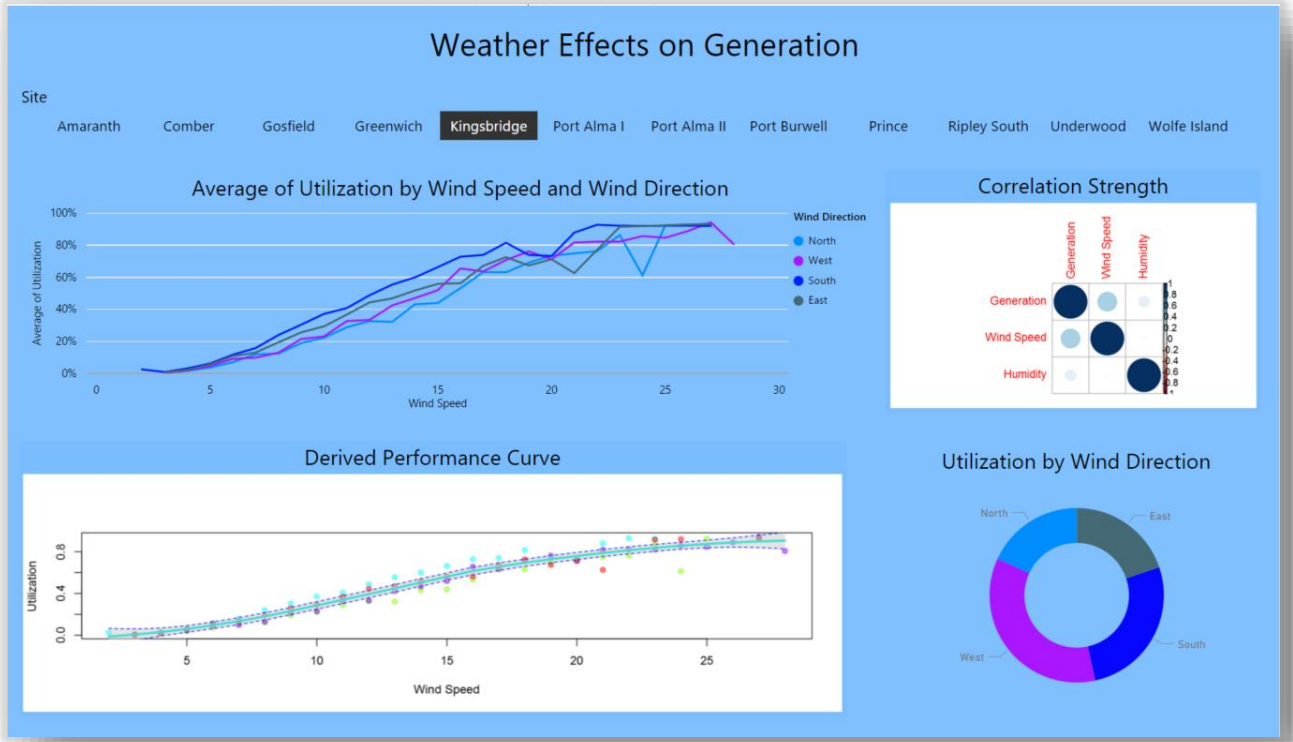
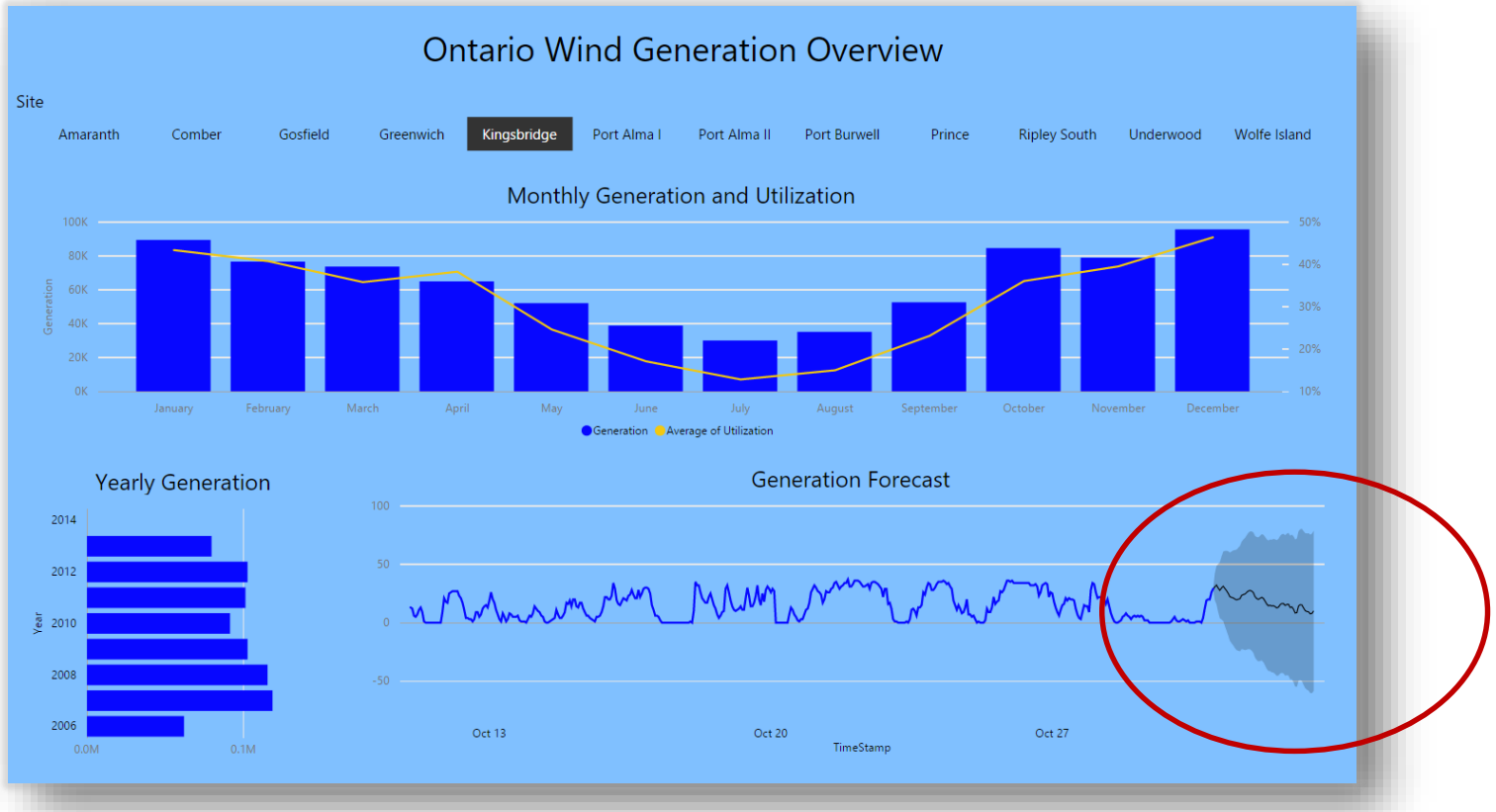
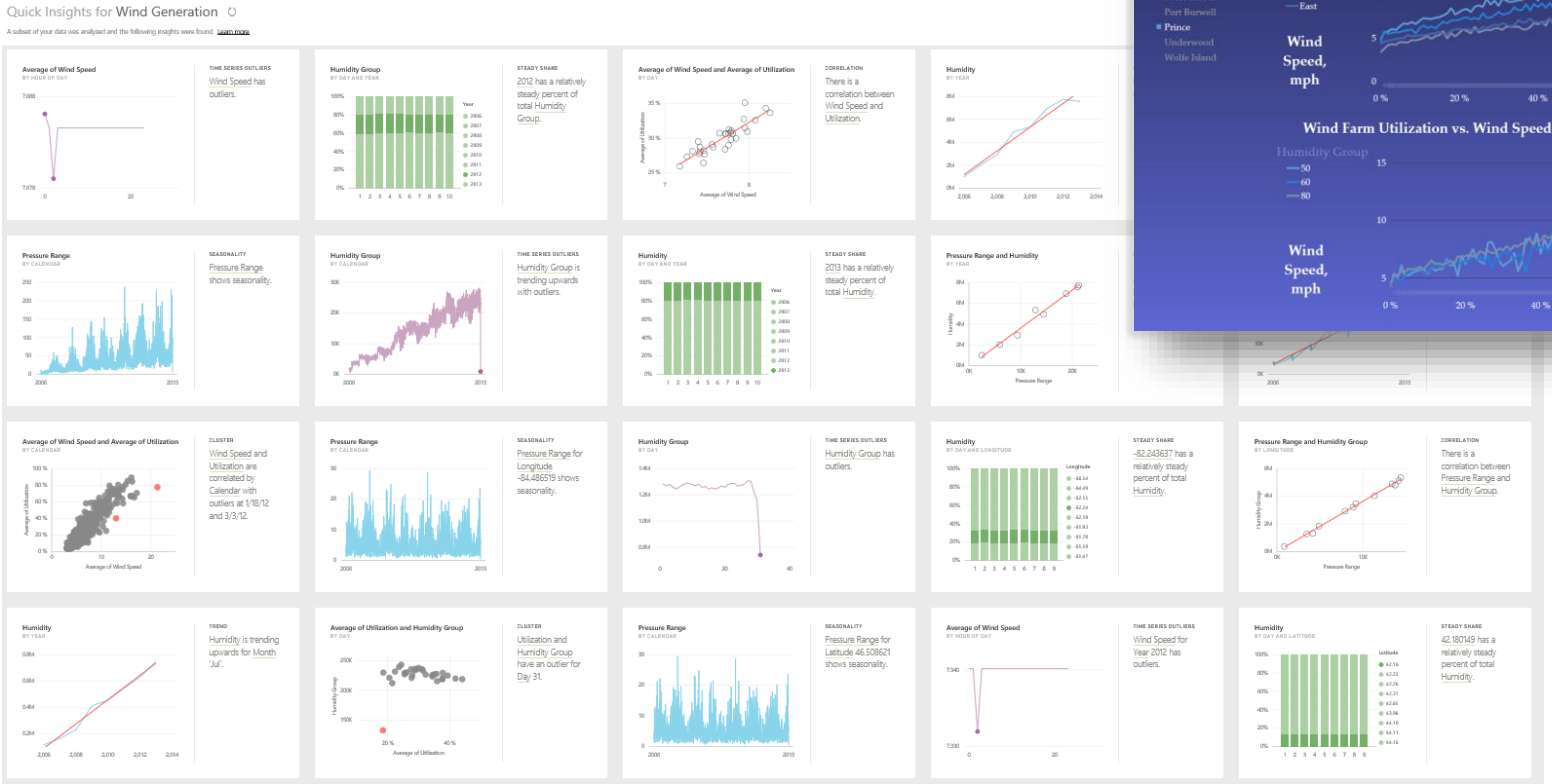
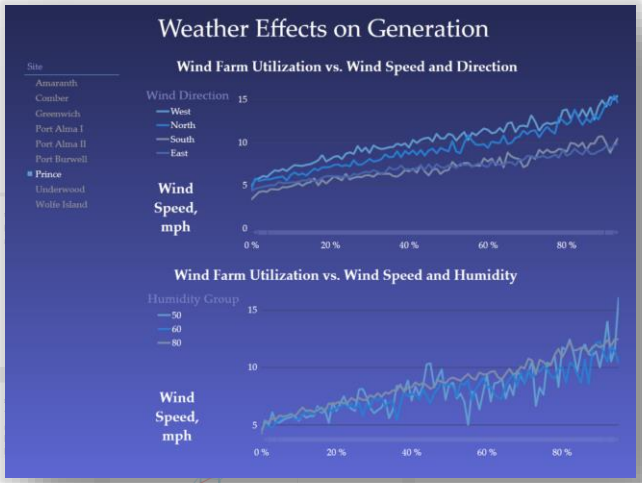


Chart-Embedded Analytics



Quick Insights for Report Data





Summary and Wrap Up

- Real time data is different. The PI System Infrastructure can supply required context, proper aggregation and configurable cleansing rules needed to prepare operational data for advanced analytics.

- Look holistically when selecting an analytical method or methods. PI Analytics is very capable of performing equation-base analytics for performance and conditions assessment, as well as, certain types of predictive analytics.

- The PI Integrator for BA establishes contextual “Common Ground” enabling organizations to leverage emerging technologies for Operational Intelligence.

Thank You

OSIsoft®