

Feeding the Machine Learning Monster

Rick Smith
Process Information Manager

International Paper

- World's Largest Pulp and Paper Company
 - Founded 1898
 - 55,000 Employees
 - \$21.7 Billion Net Sales (2017)
- OSIssoft Installed Base
 - 37 Facilities (62 PI Servers)
 - US, Brazil, Canada, France, India, Poland, Russia
 - 1.9+ Million Active PI Tags



Machine Learning Proof of Concept

- 2015 – Projects Prove Insight & Money in the Data
- 2016 – Began Enterprise Agreement
- 2017 – Mill of the Future -> Advanced Analytics Pilot
 - Test 4-5 Tools at 3 Mills (Limited Areas) -> Picked One
- 2018 – Mill of the Future -> Advanced Analytics Extended Pilot
 - 5 Mills (All Areas) -> Prove We Can Do it Again & Fine Tune Processes

Feeding the Machine Learning Engine



International Paper Needs to Extract 5-10,000 PI Tags of 1-Minute Data to Train the Machine Learning Engine.



CHALLENGE

Initial Estimates Predicted 100+ Days to Extract the Data

- Enterprise Initiative Across 30+ Facilities

SOLUTION

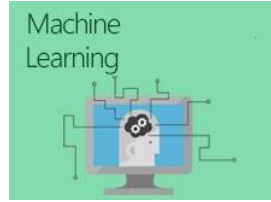
PI System tools identified largest delays and AF fixed the biggest offenders.

- PI OLEDB
- Asset Analytics
- PI Integrator for Business Analytics

RESULTS

Various techniques reduced the extraction time by 80-90%

Extracting Knowledge from Data



- **First Principles**

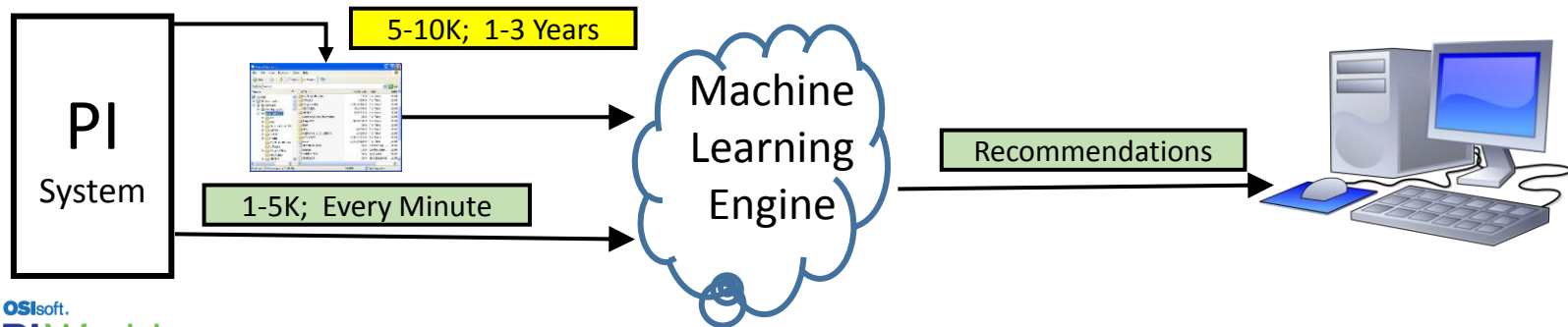
- Engineers Use Process Knowledge to Decide Important Data to Retrieve

- **Correlation Using Statistical Packages**

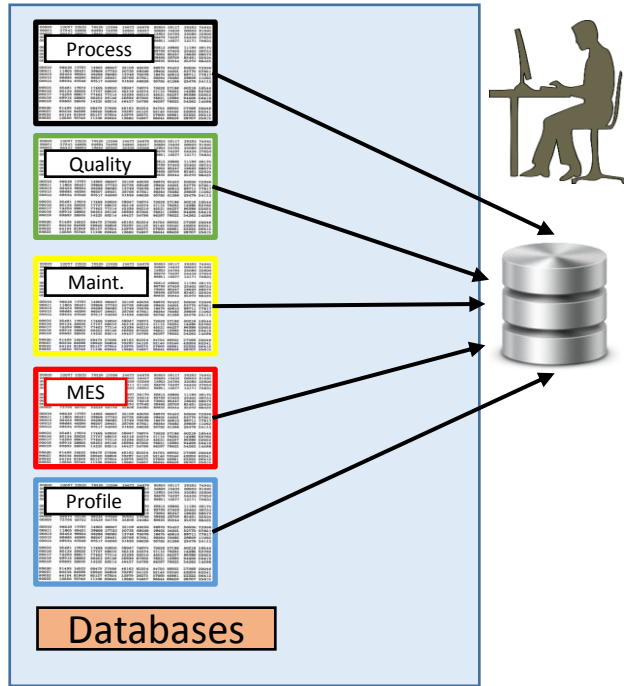
- Statistics Tells Us What is Important from Lots of Data in an Area

- **Machine Learning**

- Collection of Techniques Where Computers Learn Patterns from Data
 - History: 1-3 Years of 1-Minute Data for 5,000-10,000 PI Tags
 - Real-Time: 1-Minute Snapshots for 1,000-5,000 PI Tags



What Questions Do We Want to Ask?



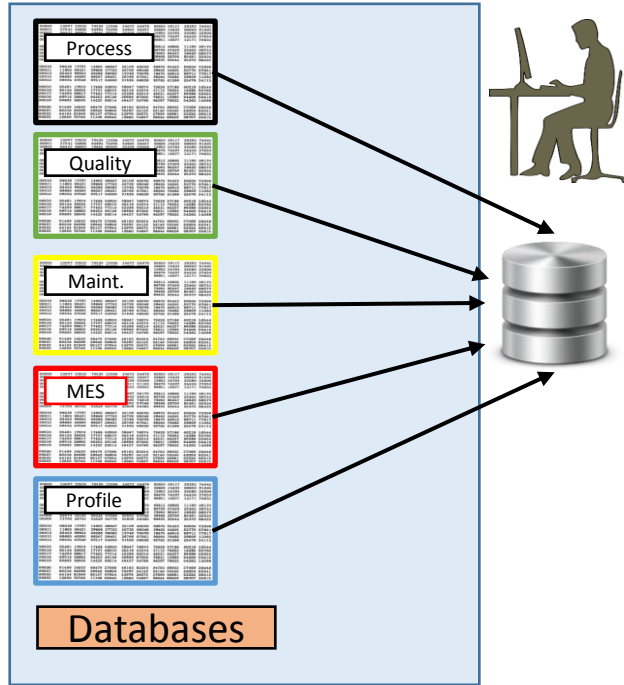
- **Old School...Reporting on Past**

- How Did Facility Perform Last Month?
- How Did Facility Do Yesterday?

- **Going Forward... Learn from Past**

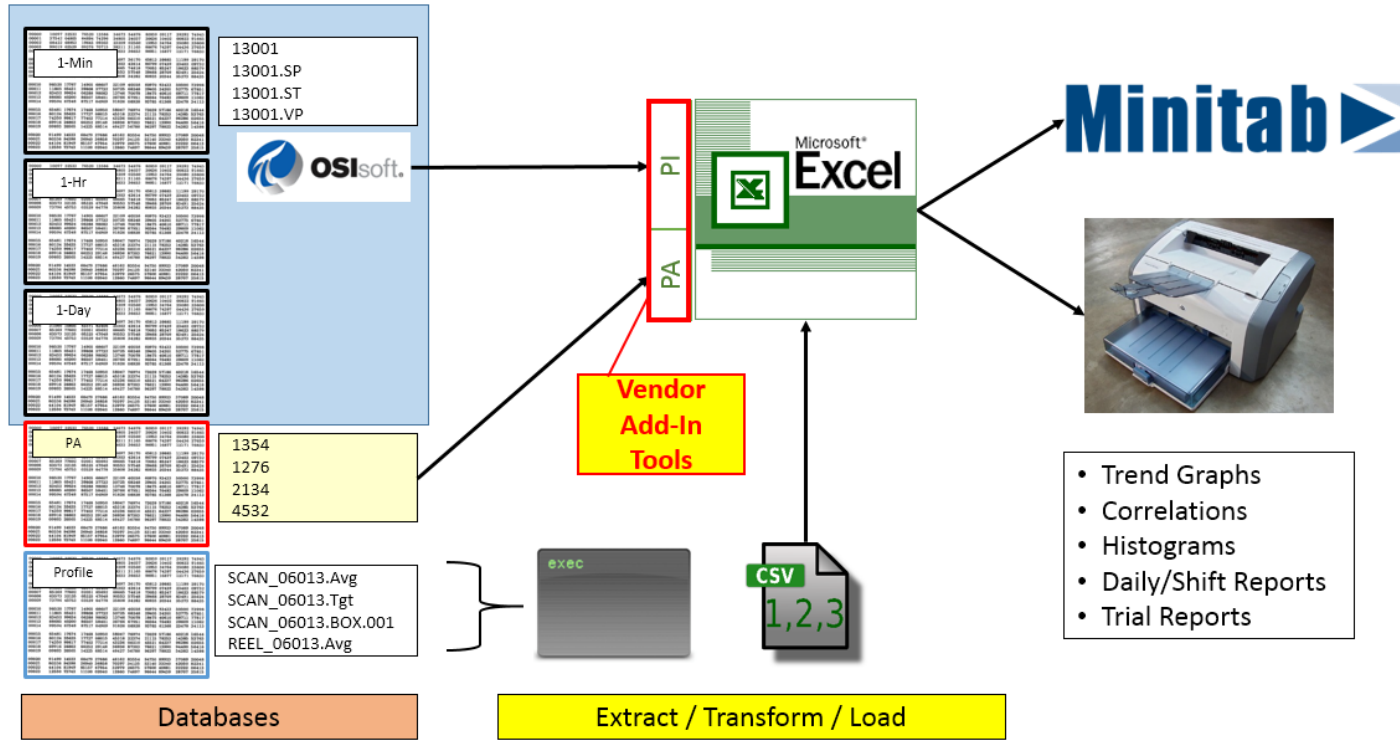
- What Operational Changes Should I Make Today?
- What Should I Do Differently Now?
- What Should I Do Differently in the Future?

What Data Do We Have?



- Business/Process Data Silos
 - **Process** – 0.25s-1min
 - **Quality** – Reel Lab Tests & Hourly Labs
 - **Profile** – 120-1000 Boxes/Sensor @ 20s-60s
 - **Maintenance** – Equipment Specs, Maintenance Records, etc.
 - **MES** – Orders, Recipes, Specifications
 - **ERP** – Feeds from Above Plus Everything Else
- We Spend 80-90% of Our Time Merging, Cleaning Data for Analysis
- **Need Good Data at the Right Frequency**

Data Collection Before 2015

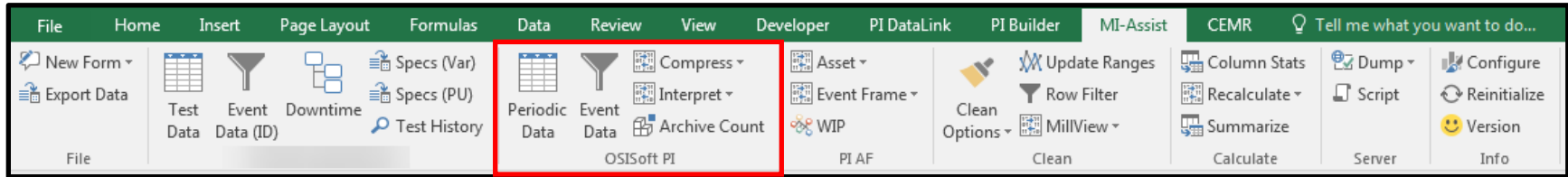


Process Audits -> Long Data Retrievals

- Retrieve 1-Year of 1-Hour Values for 50-200 Variables
- PI DataLink Issues (2015 or Earlier)
 - Data Retrieval Takes 3-12 Hours
 - “Too Many Events” Error
 - PI Server Timeouts
 - Editing Calculations Causes Datalink Refresh
 - Copy/Paste Datalink Results to Separate Sheet for Calculations
- PI DataLink (2016 and Later)
 - Performance Improved, But Still Need Alternate Solution

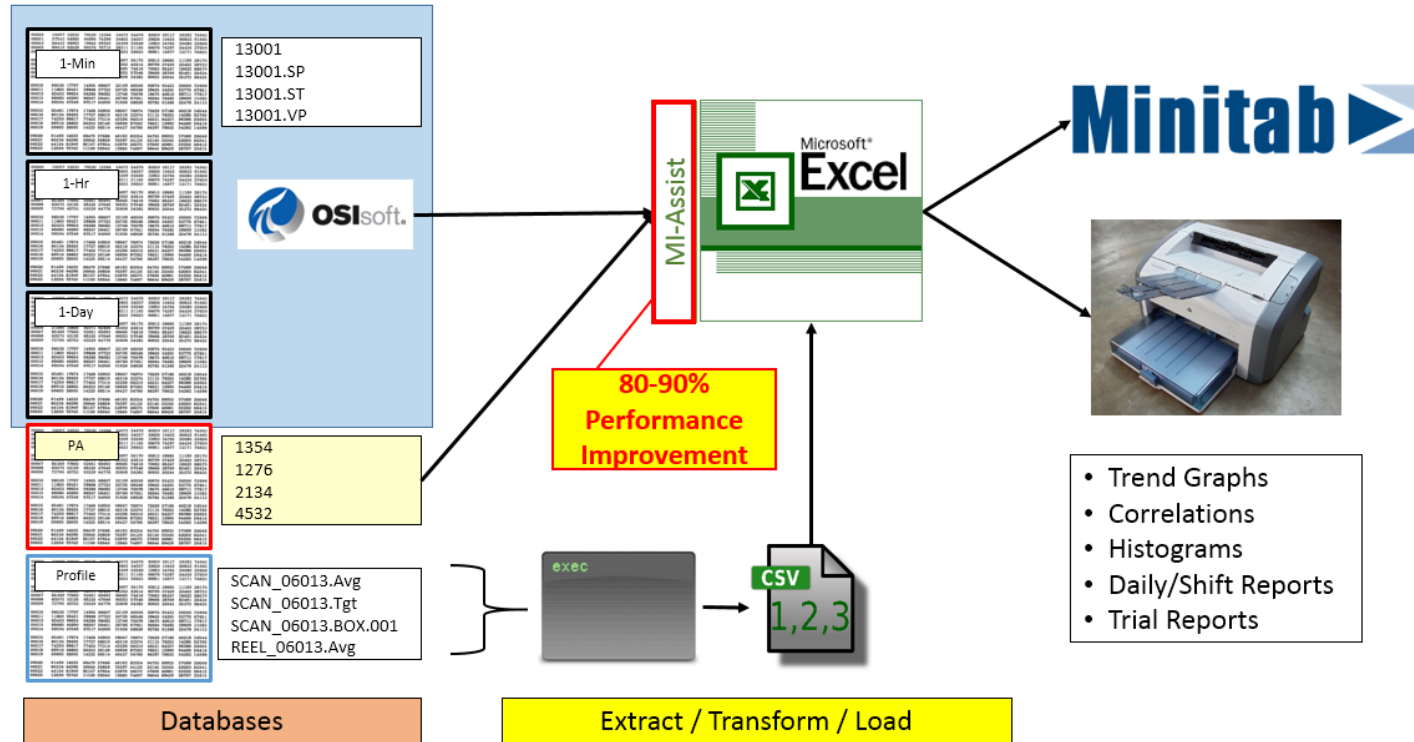


MI-Assist Addresses Process Audit Issues



- MI-Assist Excel Add-in -> PI OLEDB Provider
 - **User Requests -> PI SQL Queries -> Populate Excel**
 - Chunk Requests to Avoid Timeouts and “Too Many Events” Error
 - No Formulas to Eliminate without Copy/Paste Requirement
 - Pick-up Data Retrieval After Data Already Retrieved
 - Data Retrieval Takes 15-45 Minutes

Ad-Hoc Data Analysis After 2015 (MI-Assist)



Seconds/Tag-Year...Minutes/Tag-Year

Tag-Year /tag yir/ *noun*

The Amount of Time to Read One (1) Year of Data for One (1) PI Tag

PI Tags Read	1-Year Read Time	Seconds/Tag-Year	10K Tags / 1-Year
50	3-Hour	216.0	25 Days
100	12-Hour	432.0	50 Days
50	15-Min	18.0	2.0 Days
200	45-Min	13.5	1.6 Days

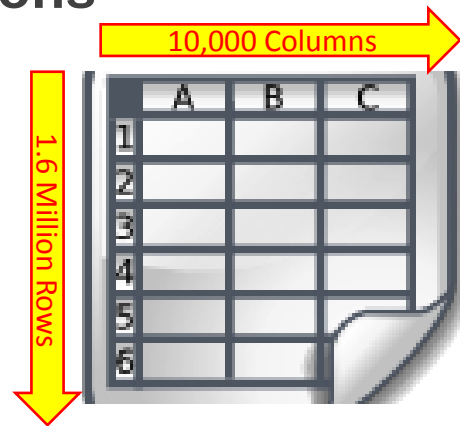
Machine Learning -> Let's Collect Big Data

- **First Machine Learning Data Request is:**

- 12-36 Months of 1-Minute Interpolated Values for 10K PI Tags
 - $3 * 365 * 1440 * 10,000 = 16 \text{ Billion Rows}$ (PI Tag, TS, Value, UOM)
- Extrapolated Test Read -> **100+ Days** (860+ Sec/Tag-Year)

- **Time to Engineer a Solution and Evaluate Options**

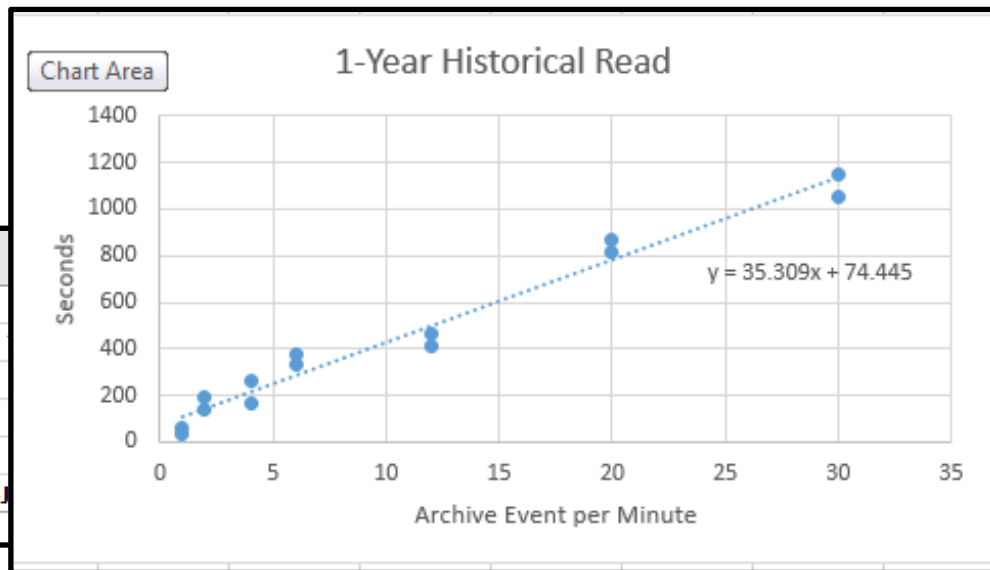
1. Use Vendor's Retrieval Mechanism
2. PI DataLink
3. PI OLEDB Provider (Tag-Based Retrievals)
4. PI OLEDB Enterprise (Asset-Based Retrievals)
5. Reprocess the Data Archives
6. Use AF to Lower Frequency of Dense PI Tags
7. PI Integrator for Business Analytics



Mill-X Archive Writes/Day (1/1/18-1/3/18)

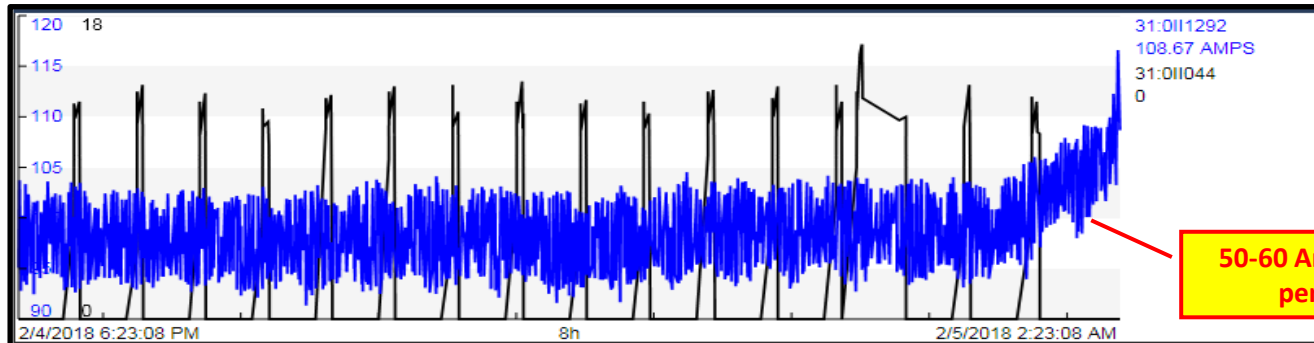
- 45,562 PI Tags; ~70 Million Archive Writes/Day
- 335 PI Tags < 2s Archive Rate -> 0.7% of PI Tags and 36.8% of Archive Writes
- 647 PI Tags < 4s Archive Rate -> 1.4% of PI Tags and 50.6% of Archive Writes
- 1134 PI Tags < 10s Archive Rate -> 2.3% of PI Tags and 58.9% of Archive Writes
- Why Does it Matter?
 - Read 1-Year (Archive 60s) -> 0.9 min
 - Read 1-Year (Archive 30s) -> 2.8 min
 - Read 1-Year (Archive 10s) -> 6.0 min
 - Read 1-Year (Archive 2s) -> 18 min

E	F	G
	Number Tags	45,562
	Total Events	68,337,418
	Average Events/Tag	1,500
3s	>28800	1.20%
10s	>8640	2.32%
pointtyl	pointsource	01-Jan-2018
D	B	0



Timing Study #1 – “Dense” PI Tags

Tag	36 sec/Tag-Yr	122 sec/Tag-Yr	137 sec/Tag-Yr	137 sec/Tag-Yr	94 sec/Tag-Yr
Timing_Row					
Source Type					
Value Type					
Eng Units					
DescriptioTimestamp					
	 Spor...freq Amps	 PM1 NO...MARY SCREEN AMPS	 PM1 MID...MARY SCREEN AMPS	 PM1 SO...MARY SCREEN AMPS	 PM1 15...uum Pump Motor



Repeat Timing Study #1 – AF Element

RM-PM1-PoorlyCompressedTags			
General Child Elements Attributes Ports Analyses Notification Rules Version			
Filter			
	Name	Unit O...	Settings...
☒	PM1 15/16 Vacuum Pump Motor	ampere	\\RM_sab4pip1\31:0II1294
☒	PM1 MIDDLE PRIMARY SCREEN ...	ampere	\\RM_sab4pip1\31:0II1292
☒	PM1 NORTH PRIMARY SCREEN A...	ampere	\\RM_sab4pip1\31:0II1291
☒	PM1 SOUTH PRIMARY SCREEN A...	ampere	\\RM_sab4pip1\31:0II1293
☒	Spool Starter Freq Amps	ampere	\\RM_sab4pip1\31:0II044

- Tag-Based -> Single Tag for Time Range
- Asset-Based -> Query All Tags for Time Range
- AF Retrieval 2-5x Faster

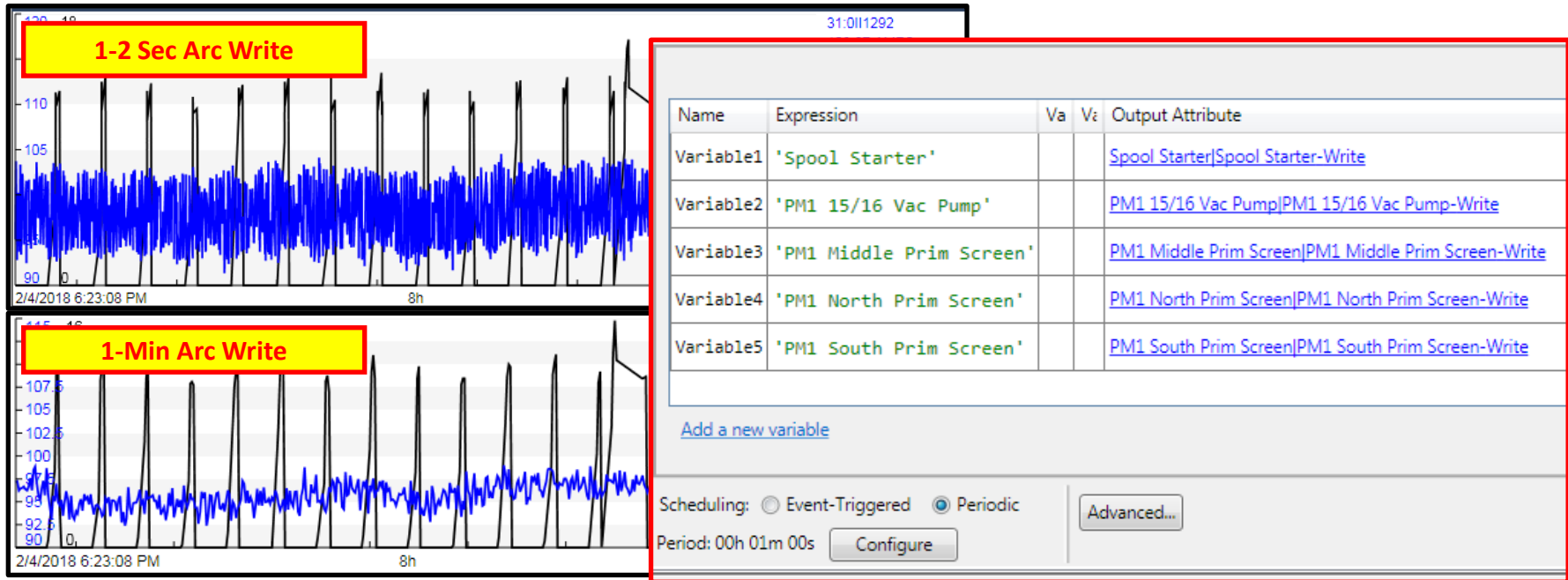
AFConfig	S:Loveland_Dev;DB:Testing-Rick Smith;ELT:EL-RM-PM1-PoorlyCompressedTags					
ELName	*					
Start Date	09-01-2017 00:00:00					
End Date	12-01-2017 00:00:00					
Frequency	1m					
Eng Units						
Description	Time	PM1 15/16 Vacuum Pump Motor	PM1 MIDDLE PRIMARY SCREEN AMPS	PM1 NORTH PRIMARY SCREEN AMPS	PM1 SOUTH PRIMARY SCREEN AMPS	Spool Starter Freq Amps

v2017.12.07 -> PIAF RunTime: 04:34; PI RunTime:10:42 (Office)

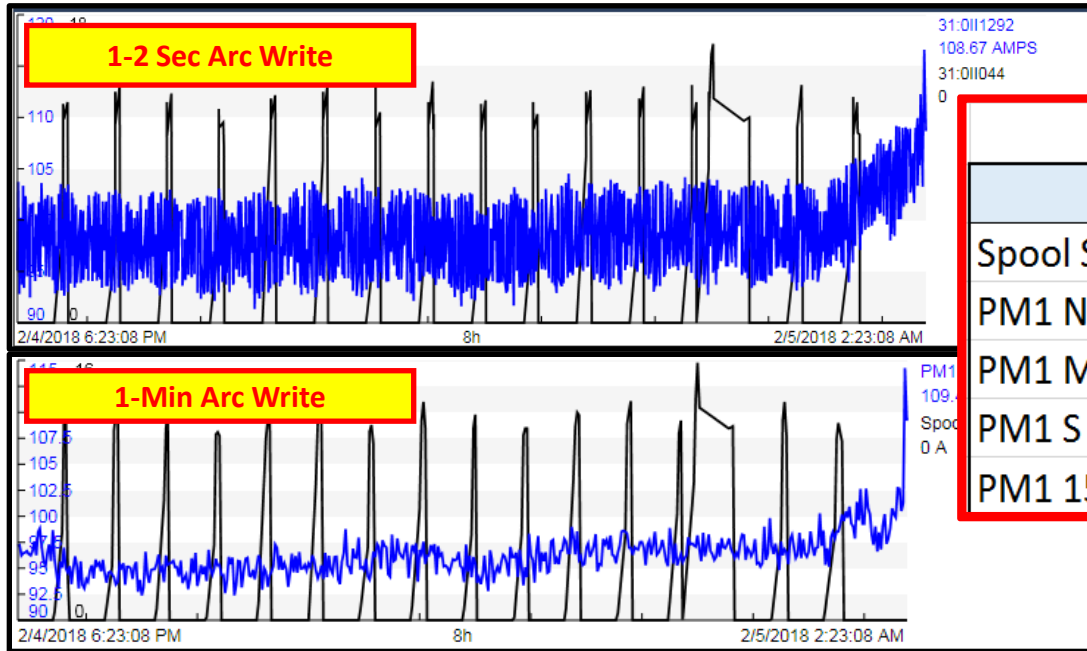
v2017.12.07 -> PIAF RunTime: 03:09; PI RunTime:08:26 (VPN)

v2017.12.08-> PIAF RunTime: 01:20; PI RunTime:02:15 (Office)

TS #1 - Reduce 1s Archives Writes to 1m



TS #1 - Reduce 1s Archives Writes to 1m



	3-Month Read	
Description	Raw	1-Min
Spool Starter Amps	0.15	0.100
PM1 N Prim Scrn Amps	0.51	0.079
PM1 Mid Prim Scrn Amps	0.57	0.079
PM1 S Prim Scrn Amps	0.57	0.094
PM1 15/16 Vac Pmp Amps	0.39	0.080

85% Decrease in Read Time

Timing Study #1 Conclusions

- Archive Write Count Correlates to Data Retrieval Time
 - AF Element Retrievals Better Than Single PI Tag Reads
 - Reducing Data Density Reduces Data Retrieval Time
- ... Let's Try Something Bigger...

[illegible][illegible]

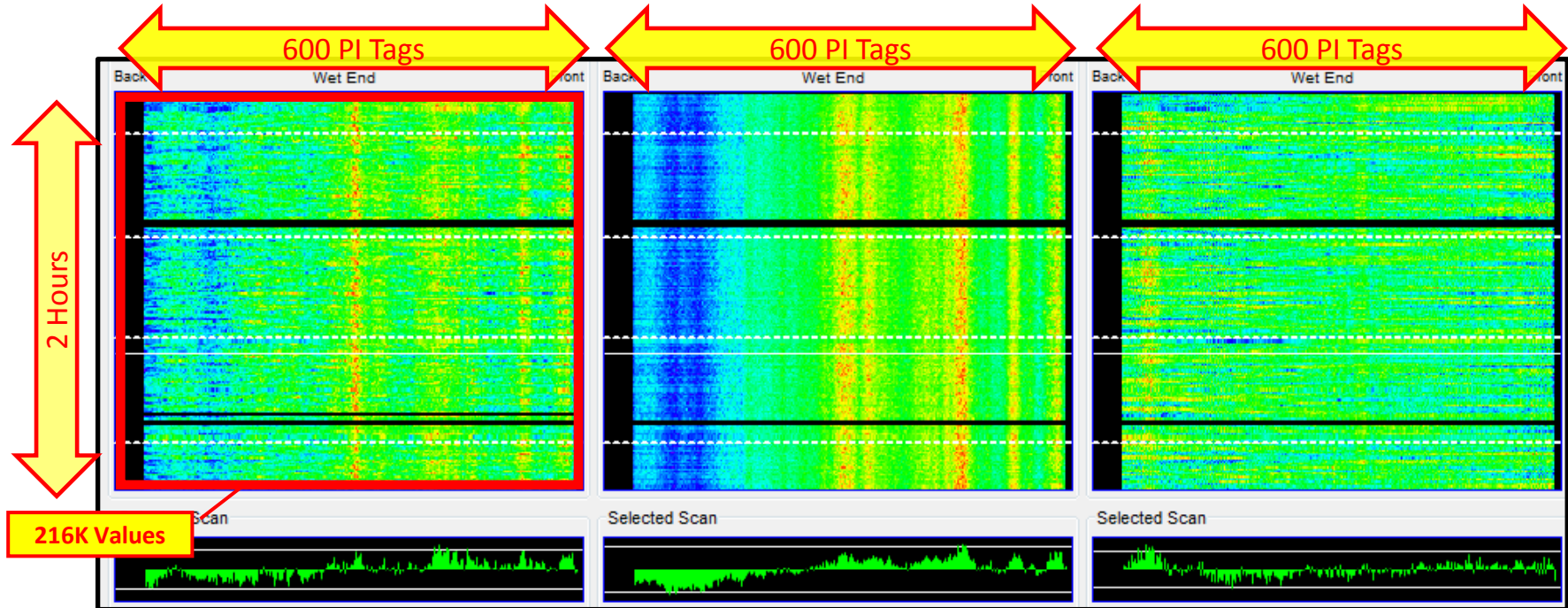
Timing Study #2 – 50 PI Tags (Audit Simulation)

Start Date	12-01-2016 00:00:00									
End Date	12-01-2017 00:00:00									
Minutes per PI Tag (1-Yr. Read)	0.41	0.25	0.23	0.24	3.29	2.31	1.29	1.00	0.58	0.58
Source Type	321LIC022	322LIC322	322LIC323	322LIC324	321FIC009	321FIC012	321FIC055	321FIC013	322FIC312	322FIC309
Value Type	PI	PI	PI	PI	PI	PI	PI	PI	PI	PI
Eng Units	SNAP	SNAP	SNAP	SNAP	SNAP	SNAP	SNAP	SNAP	SNAP	SNAP
DescriptioTimestamp	RPM	FT	FT	FT	GPM	GPM	GPM	GPM	GPM	GPM
	#1 LINE W	#2 LINE W	#2 LINE W	#2 LINE W	#1 LINE W	#1 LINE W	ENV - A LI	#1 LINE W	L2 WSHR1	#2 LINE W



AFConnig	S:\Cleveland_Dev\DB:Testing-Rick Smith;ELT:EL-AFwideDump														
AFELEMENTName	RR-Buckman1														04:52 (50 AF Attributes)
Start Date	12-01-2016 00:00:00														
End Date	12-01-2017 00:00:00														
Frequency	1m														3x Improvement
Eng Units															
Description	Time	Tag01	Tag02	Tag03	Tag04	Tag05	Tag06	Tag07	Tag08	Tag09	Tag10	Tag11	Tag12	Tag13	

Timing Study #3 - Gauge Profile Tags (AF)



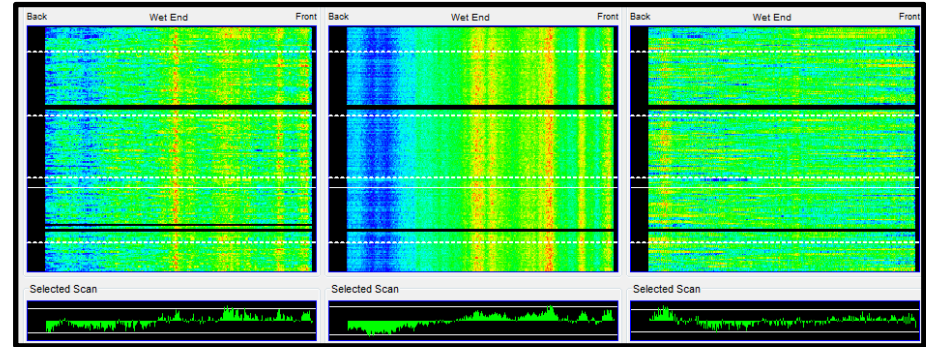
Timing Study #3 - Gauge Profile Tags (AF)

- Retrieve 1-Minute Profile Summary

- Ten (10) Sensors (40 Values)
 - 2-3 Profile Scans/Minute
- 8-Hr -> 5 Seconds
- 14-Days -> 40 Seconds
- 25 sec/Tag-Year

- Retrieve 1-Minute Full Profile

- Ten (10) Sensors (6,040 Values)
- 8-Hr -> 02:00
- 1-Day -> 06:30
- 15 sec/Tag-Year



Timing Study #4 – Reprocess Archives

- Select 100 PI Tags
- Read 3-Months of 1-Minute Data
 - 65 Minutes
- Reprocess the Archives
- Read 3-Months of 1-Minute Data
 - 28 Minutes... 50% Reduction



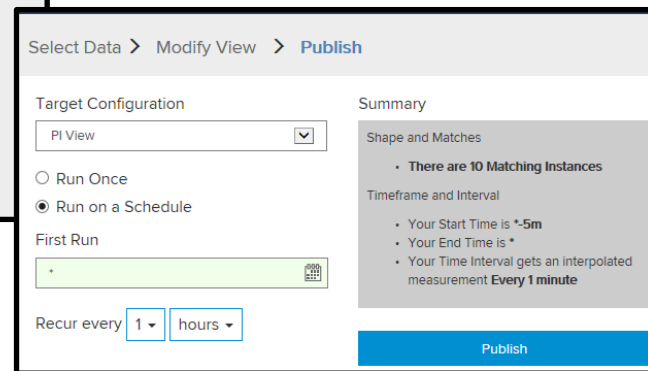
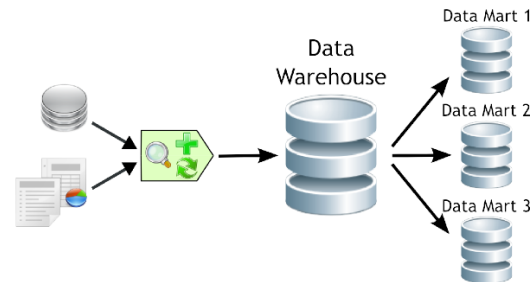
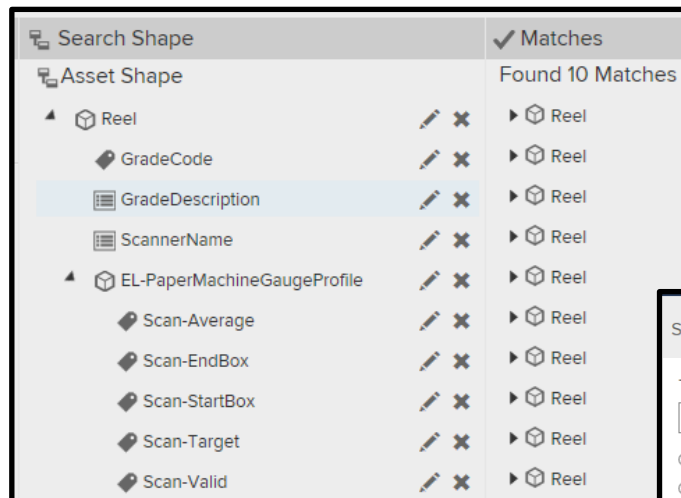
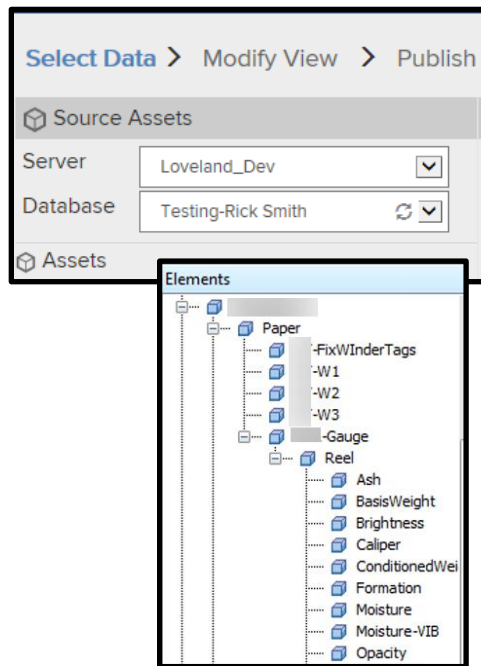
Good Progress... What Else Can We Do?



1. Create AF Elements for Data Retrieval ✓
2. Create 1-Minute Machine Learning PI Tags w/ AF ✓
3. Reprocess the Data Archives ✓
4. Create a Data Mart at Required Frequency
 - a. PI Integrator for Business Analytics
5. Fix PI Tag Compression to Stop Storing Sensor Noise
 - a. Forward in Time Only

Create Data Mart View

PI Integrator for Business Analytics



View Data... Send to Data Mart

Select Data > **Modify View** > Publish

+

Add Column

11 columns

▼

Edit Row Filters

0 Row Filters

≡

Edit Value Mode

Interpolated Values

Every 1 minute

Start Time

*-5m

0000

End Time

*

0000

Reel	TimeStamp	GradeCode	GradeDescrip	ScannerName	EL-PaperMac	Scan-Average	Scan-EndBox	Scan-StartBo	Scan-Ta
Reel	2/6/2018 9:18:23.801...	006080200...	0	Reel	BasisWeight		576	14	92
Reel	2/6/2018 9:19:23.801...	006080200...	0	Reel	BasisWeight		576	14	92
Reel	2/6/2018 9:20:23.801...	006080200...	0	Reel	BasisWeight		576	14	92
Reel	2/6/2018 9:21:23.801...	006080200...	0	Reel	BasisWeight		576	14	92
Reel	2/6/2018 9:22:23.801...	006080200...	0	Reel	BasisWeight		576	14	92
Reel	2/6/2018 9:23:23.801...	006080200...	0	Reel	BasisWeight		576	14	92
Reel	2/6/2018 9:18:23.801...	006080200...	0	Reel	Brightness		576	14	92
Reel	2/6/2018 9:19:23.801...	006080200...	0	Reel	Brightness		576	14	92
Reel	2/6/2018 9:20:23.801...	006080200...	0	Reel	Brightness		576	14	92
Reel	2/6/2018 9:21:23.801...	006080200...	0	Reel	Brightness		576	14	92
Reel	2/6/2018 9:22:23.801...	006080200...	0	Reel	Brightness		576	14	92
Reel	2/6/2018 9:23:23.801...	006080200...	0	Reel	Brightness		576	14	92

Maybe Faster...

Runs on Another Computer...

**Maybe Faster...
Runs on Another Computer...**

Fix PI Tag Compression

- **Long-Term Problem**

- Many of Stakeholders
- Applications Need Data at Different Data Frequencies
- People Do Not Understand Noise
- Polled Interfaces Switching to OPC (Exception-Based)
- Big Data and Data Retrieval Raises Awareness of the Issue
- Gathering Stakeholders to Work Problem



Stakeholders and PI Tag Compression

- **No Compression Camp** (aka Store Everything)
 - Instrument Electrician / Process Control
 - Reliability Engineers
- **Compression is OK, But Not Too Much**
 - Process Engineers
 - Subject Matter Experts
 - Advanced Analytic Engines
- **Summarize the Data** For Me for High Performance Retrieval
 - Subject Matter Experts
 - Data Warehouse



Selecting the Best Option...For Now...



1. Create AF Elements for Data Retrieval
2. **Reprocess the Data Archives**
3. **Create 1-Minute Machine Learning PI Tags w/ AF**
 - a. **Relatively Easy (aka Fast) to do**
 - b. **Selective... Pick and Choose When to Apply**
 - c. **Does Not Affect Existing Applications and/or Data Consumers**
4. Create a Data Mart at Required Frequency
 - a. PI Integrator for Business Analytics
5. Fix PI Tag Compression to Stop Storing Sensor Noise
 - a. Longer Term Solution

Rick's Take-Aways

- Dense PI Tags **WILL** Slow Down Long-Term Data Pulls
- Dense Data PI Tags May Be Caused By...
 - OPC Interfaces Changed from Polling to Exception
 - Asset Analytics Triggering on Events (Consider Periodic)
- **Create Lower Frequency PI Tags for Dense Data...**
- PI Tag Compression Can Reduce Data Density, but at a Cost
- PI System Tools Can Find the Problems...
- PI System Tools Can Fix the Problems...
- **Systems Need Janitors Just Like They Need Architects**

Feeding the Machine Learning Engine



Rick Smith

Richard.Smithjr@ipaper.com

Process Information Manager
International Paper

Questions

Please wait for the **microphone** before asking your questions

State your **name & company**



Please remember to...

Complete the Online Survey for this session



Download the Conference App for OSISOFT Users Conference 2017

- View the latest agenda and create your own
- Meet and connect with other attendees



search **OSISOFT** in the app store

Merci

谢谢

Спасибо

Danke

Gracias

Thank You

감사합니다

ありがとう

Grazie

Obrigado

Systems Need Janitors as Much as They Need Architects