Streaming Analytics, Data Lakes and PI Integrators

Presented by Matt Ziegler
Daniele Farris

©2017 OSIsoft, LLC

Overview

©2017 OSIsoft, LLC

A singular data system that does all this and more

Reduce downtime up to 50%

Reduce capital investment by 3-5%

Eliminate 70% of breakdowns

Reduce maintenance costs by up to 25%



Reduce scheduled repairs by up to 12%

Cut unplanned outages by up to 50%

©2017 OSIsoft, LLC

Presentation Goals

- Explore common Enterprise Data Architectures including
  - Lambda & Kappa
  - Data Lakes

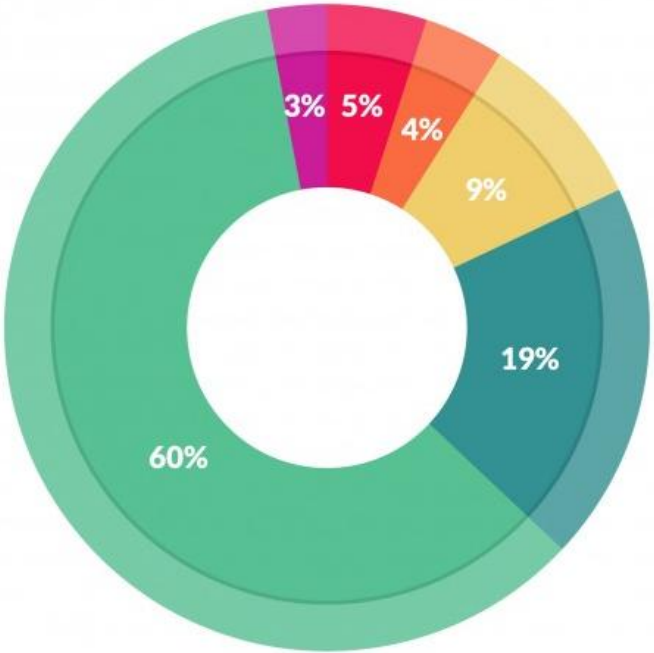
- Clarify the role the PI System plays in each architecture

- **Discuss how using the PI System helps maximize the effectiveness of scarce machine learning and analytics talent**

- Demonstrate PI Integrator for Business 2017 with Kafka in a machine learning and streaming scenario

©2017 OSIsoft, LLC

Data Scientist is the sexiest job of 21st century, but…



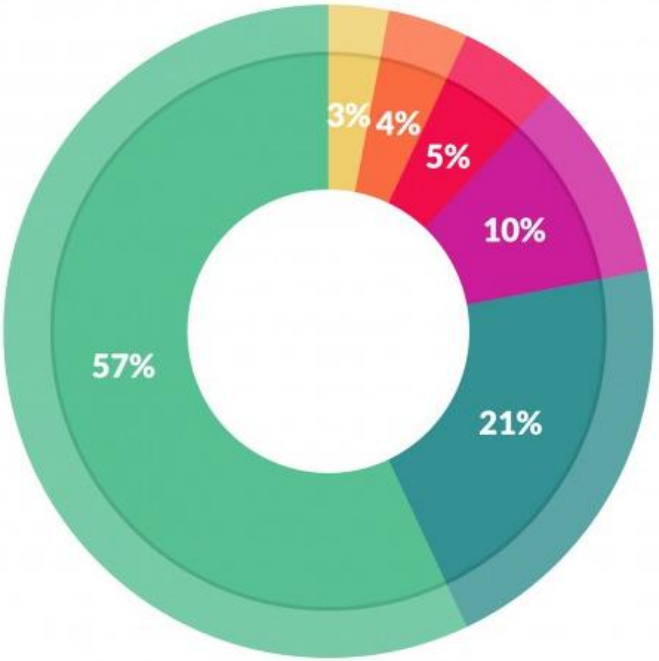
What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

Source: *http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#5481f6037f75*

©2017 OSIsoft, LLC

Data Scientist is the sexiest job of 21st century, but…

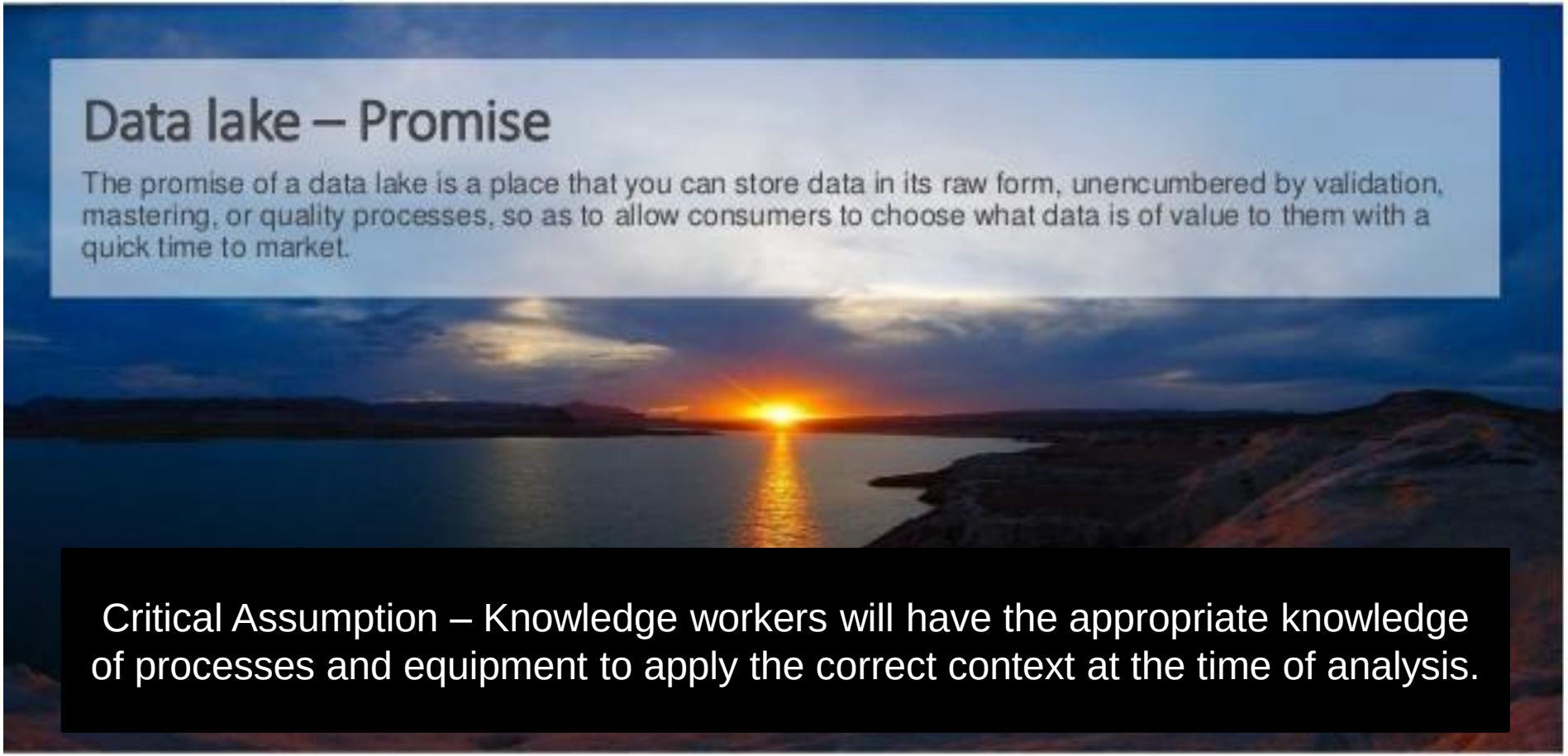


What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

Source: *http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#5481f6037f75*

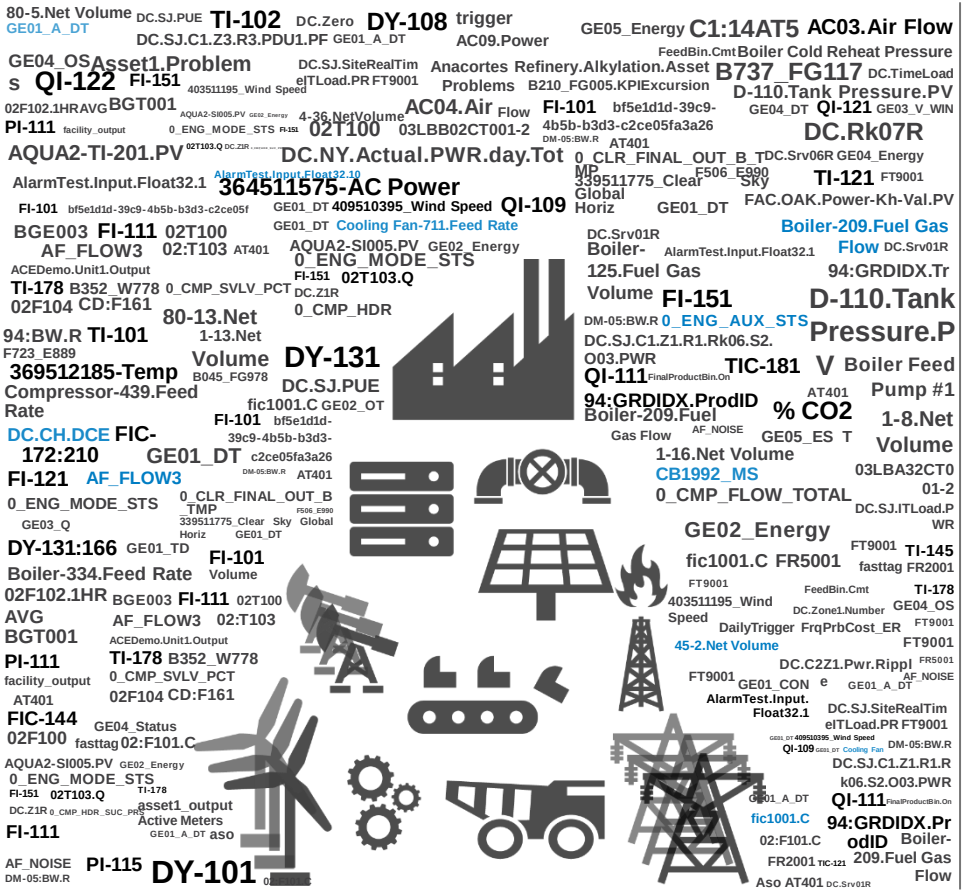
©2017 OSIsoft, LLC



NVISIA

©2017 OSIsoft, LLC

Challenge: Map raw data to the physical world



Crude Desalter

Operating Pressure: 110 psi
Charge Rate: 14 gph
Mix Valve Pressure: 8 psi
Water Rate: 8%

Weather Conditions

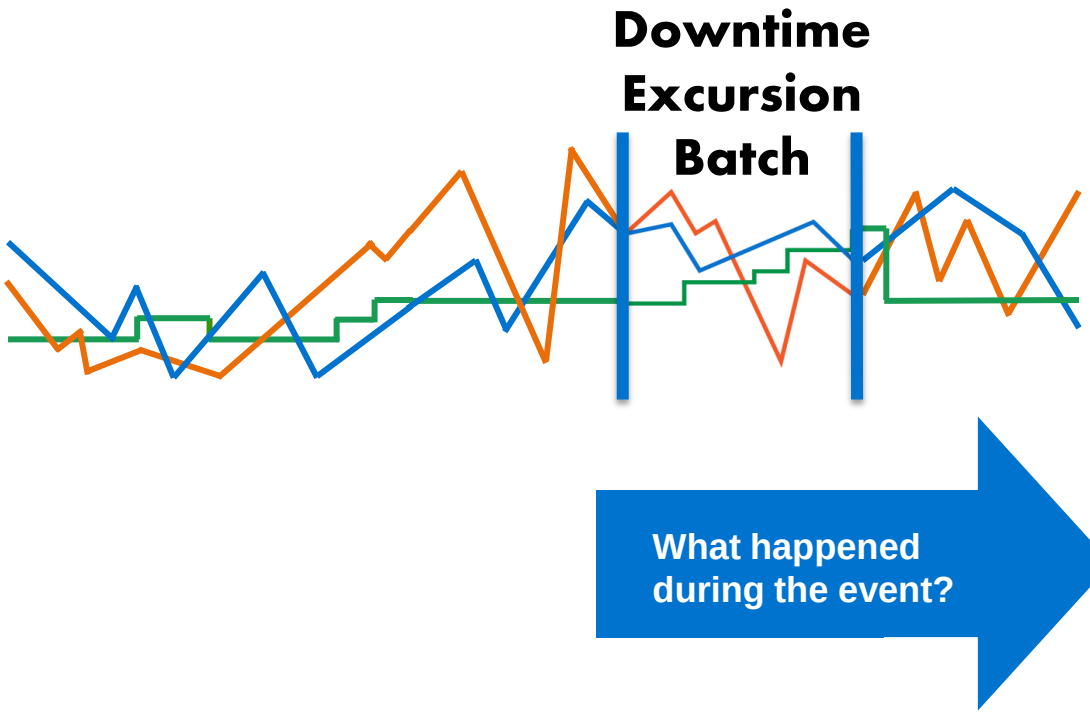
Relative Humidity: 34%
Current Temp: 85 F
High: 92
Low: 57 F
Wind: 8 mph/N

Crude Furnace

Draft Pressure: -0.5 WC
Stack Temp: 316 F
Oxygen: 2.5%
Outlet Temp: 840 F
Cold Oil Velocity: 6 ft/sec

©2017 OSIsoft, LLC

Challenge: Find events in raw data streams



Event Attribute	Value
Name	Ex 20151215-0002
Start	15-Dec-2016 10:35:02
End	15-Dec-2016 11:47:26
Asset	Boiler-352
Excursion Type	High Violation
Fuel Gas Flow.Avg	37.12 k sft3/h
myKPI.Max	47.19 bbl/d

©2017 OSIsoft, LLC

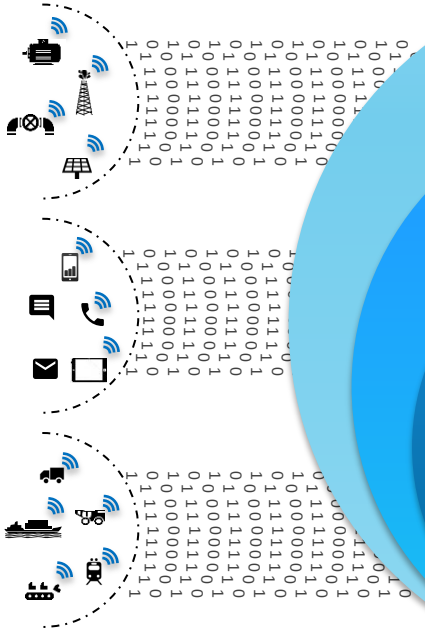
PI System collects, organizes, and enriches data

PI Interfaces & PI Connectors

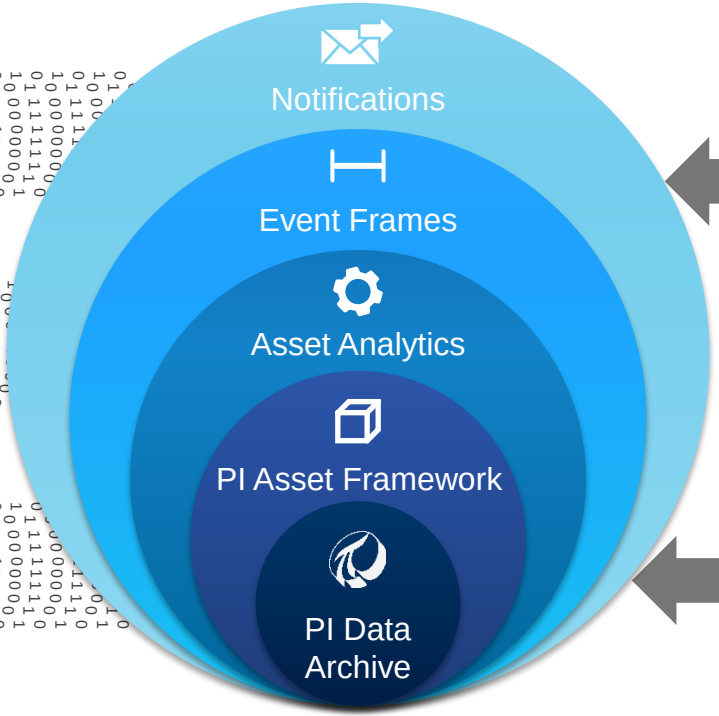
Automation or Control Systems

New Sensor Technology

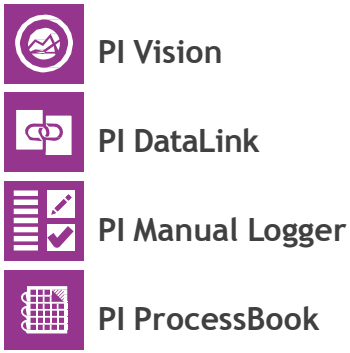
Remote and Mobile Assets



PI Server



PI Visualization Suite



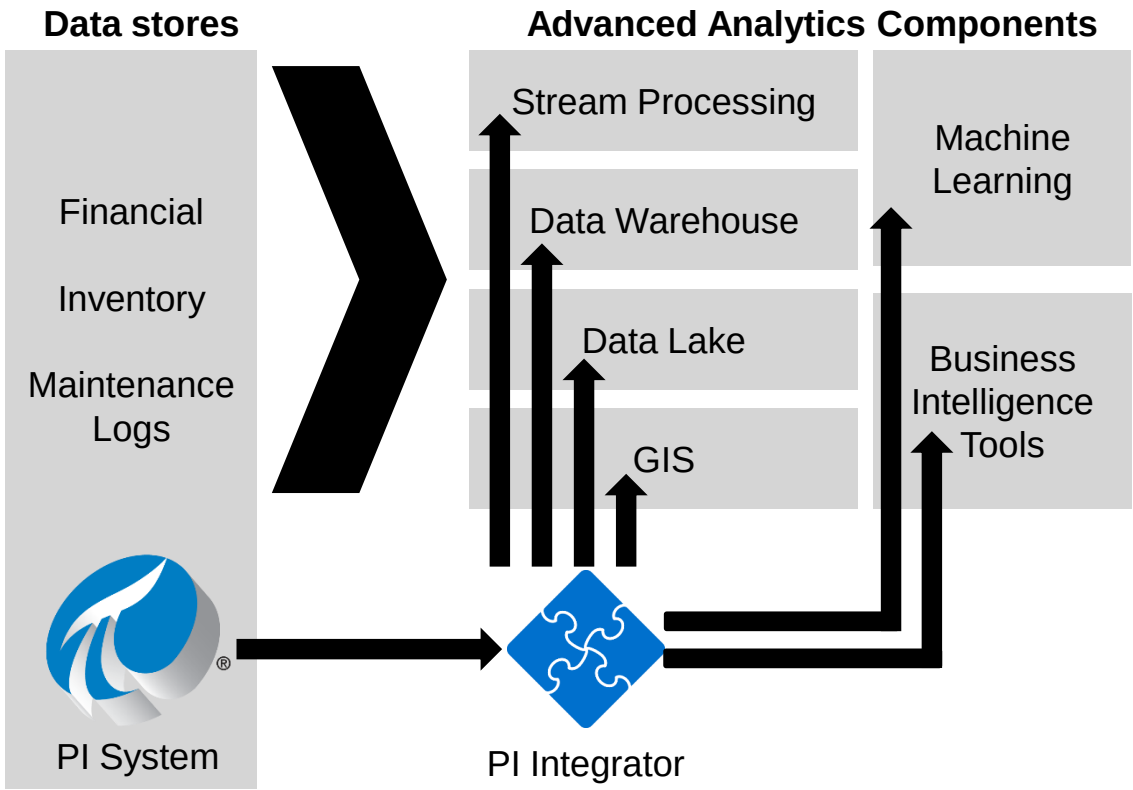
PI System Access

SQL (OLEDB, JDBC, ODBC)
Web API (REST)
OPC (DA, HDA)
SDK (.NET-based)

©2017 OSIsoft, LLC

Challenge: Bring trustworthy data to advanced analytics

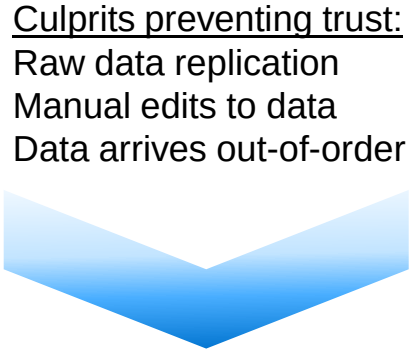
PI System can **integrate** across
advanced analytics architectures
and
deliver **authoritative** operational data



©2017 OSIsoft, LLC

Challenge: Bring trustworthy data to advanced analytics

<u>Culprits preventing trust:</u>
Raw data replication
Manual edits to data
Data arrives out-of-order



<u>Impact:</u>
Unrealistic predictions
Two databases not synchronized
Analytics disconnected from physics
Incorrect calculations

Data stores

- Operational
- Financial
- Inventory
- Maintenance Logs

Advanced Analytics Components

- Stream Processing
- Data Warehouse
- Data Lake
- GIS
- Machine Learning
- Business Intelligence Tools

Combine data to:

- Report
- Slice and dice
- Predict
- Train algorithms

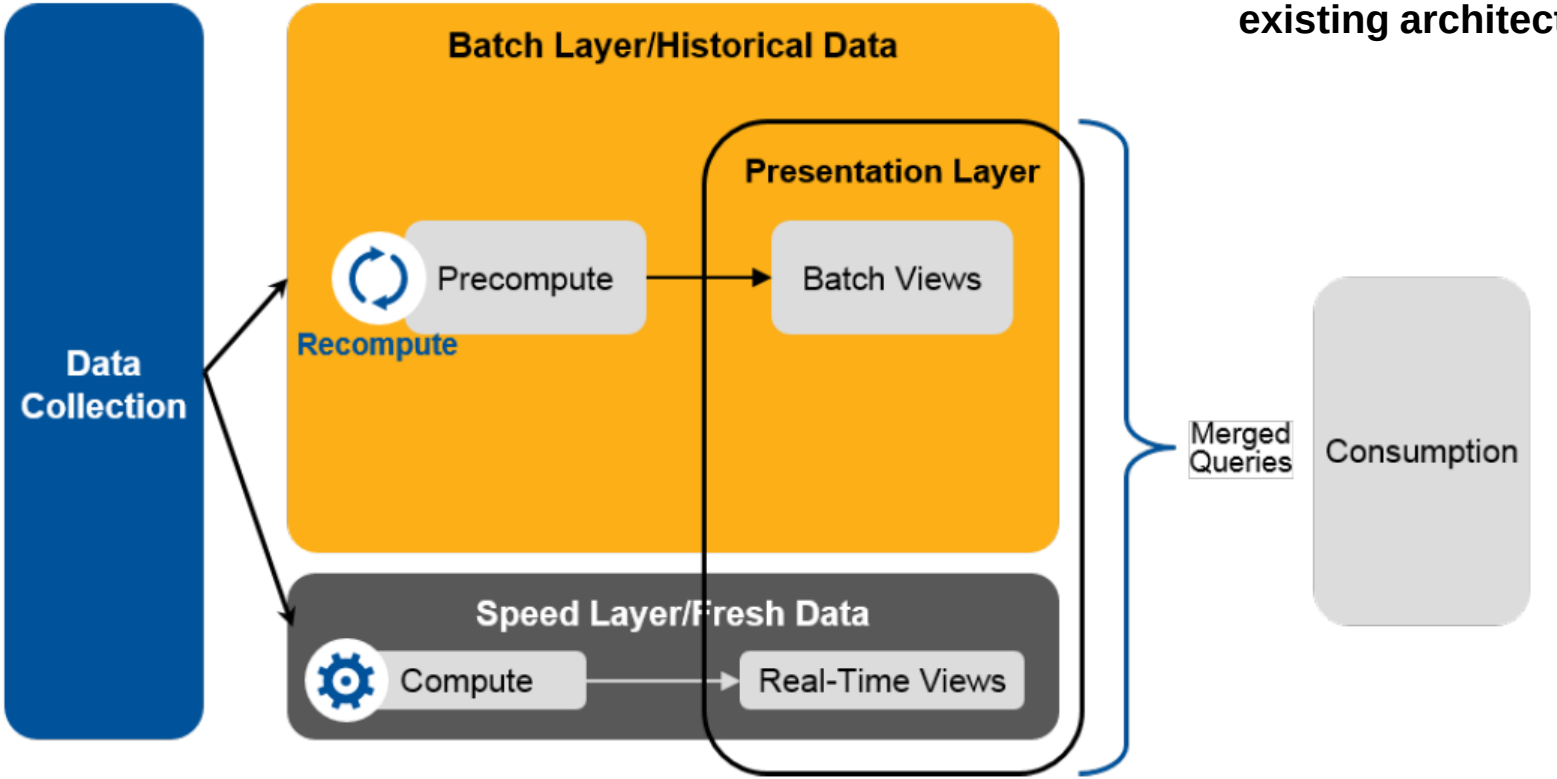
©2017 OSIsoft, LLC

Architectures

©2017 OSIsoft, LLC

Figure 1. Lambda Architecture

Add speed and scale to existing architectures

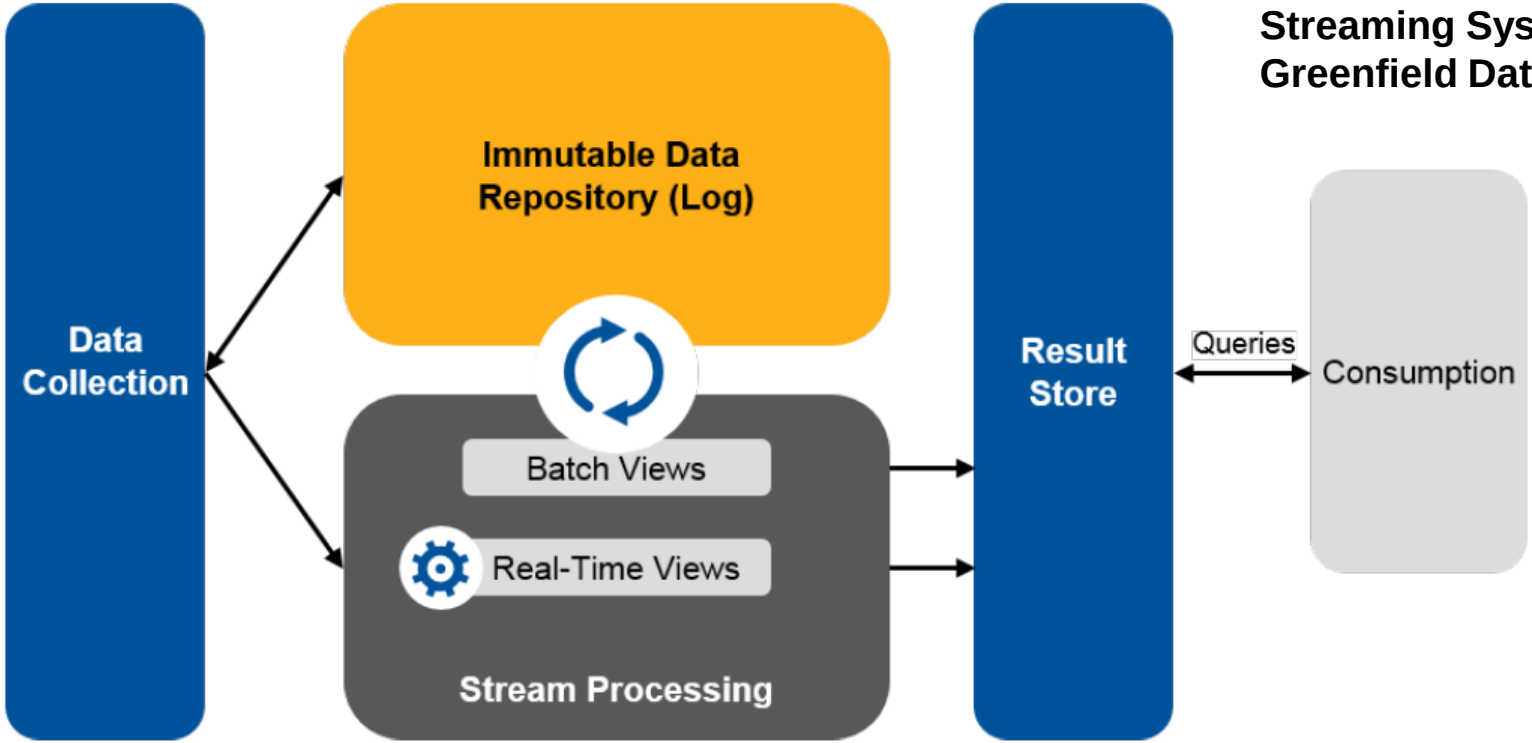


Source: Gartner (November 2016)

©2017 OSIsoft, LLC

Figure 2. Kappa Architecture

Reference design pattern for Industrial IoT Applications and Streaming Systems and Greenfield Data Systems

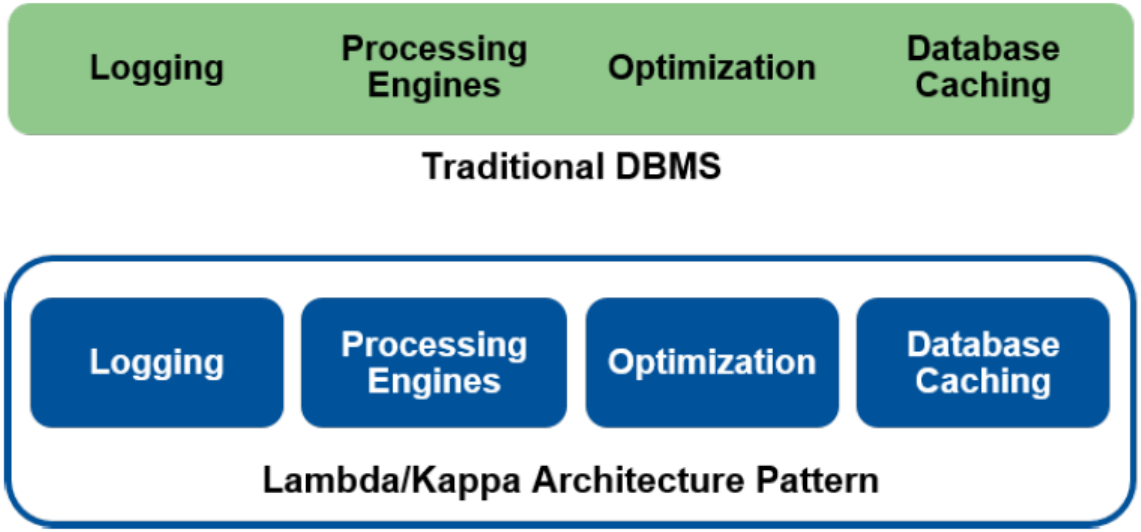


Source: Gartner (November 2016)

©2017 OSIsoft, LLC

Shift from single vendor integrated solutions to best of breed customer integrated solutions

Figure 3. Lambda/Kappa Architecture as a DBMS Implementation Pattern



A Single Product

A Set of Tools

Build Your Own DBMS

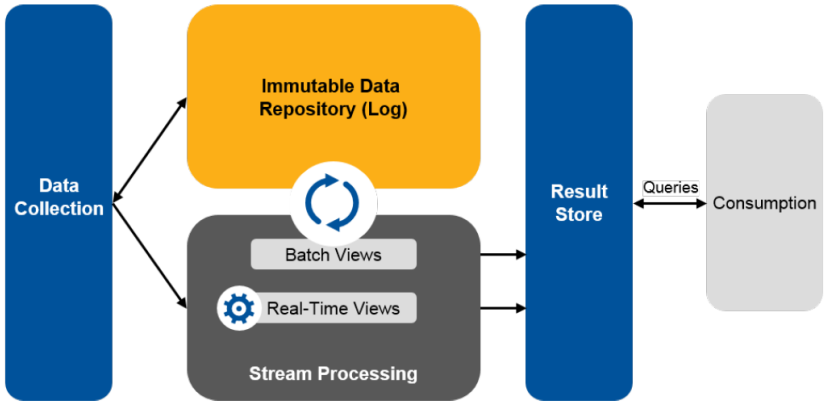
Source: Gartner (November 2016)

Be aware of the transfer of risk from vendors to open source, IT, SIs

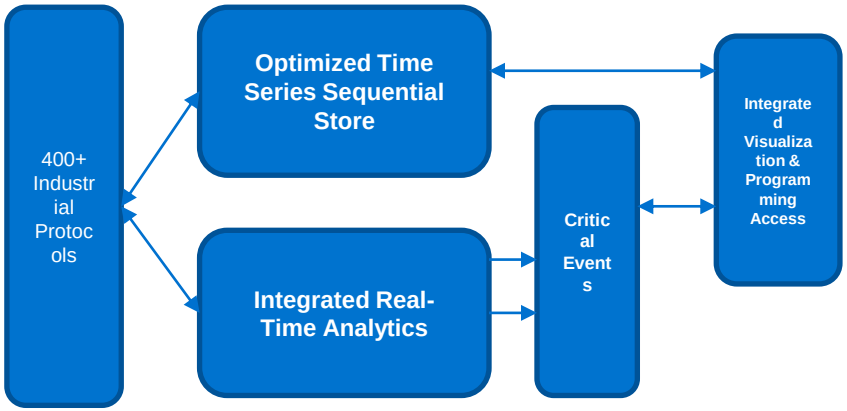
©2017 OSIsoft, LLC

PI System – An Integrated Kappa Architecture System

Figure 2. Kappa Architecture

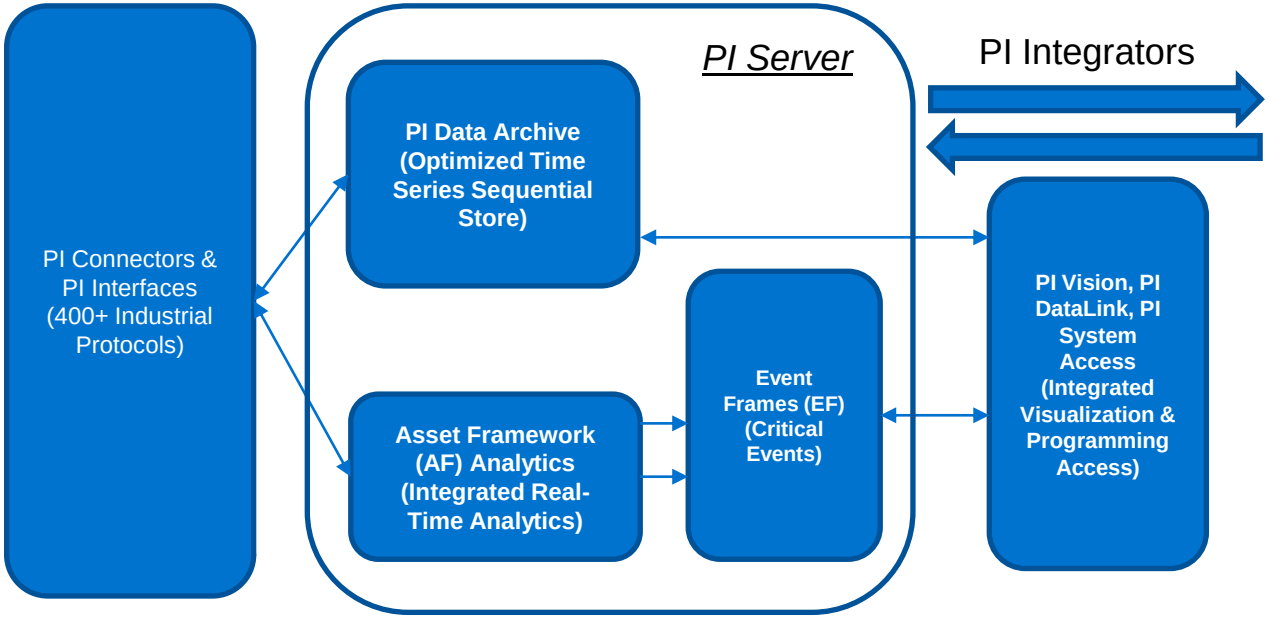


Source: Gartner (November 2016)



©2017 OSIsoft, LLC

Product Portfolio (Infrastructure & Advanced Integrations)



External Systems

kafka

Microsoft SQL Server

PI Integrator for Business Analytics

ORACLE

Spark

SAP

PI Integrator for SAP HANA

PI Integrator for Microsoft Azure

©2017 OSIsoft, LLC

Roadmap & Positioning

©2017 OSIsoft, LLC

*PI Integrators **augment** the existing set of **PI** capabilities like visualization, data access, and analytics with capabilities that **make it easy** to **interact** with non-OSIsoft **tool sets**.*

©2017 OSIsoft, LLC

PI Integrators – Climbing the Value Ladder

Disparate assets or interacting one-by-one

Interacting with common assets as a fleet

Complexity

Monitoring

Real-time visibility



- HMI

Process Optimization

Real-time & historical view across plant asset



- PI ProcessBook
- PI Coresight
- PI Datalink

Benchmarking

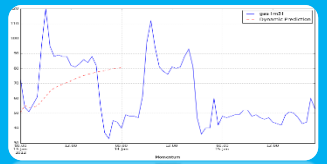
Fleet-wide performance comparison



- BI App (i.e. Tableau, Spotfire, Lumira)
- PI Integrator for Business Analytics
- PI Integrator for SAP HANA

System Optimization

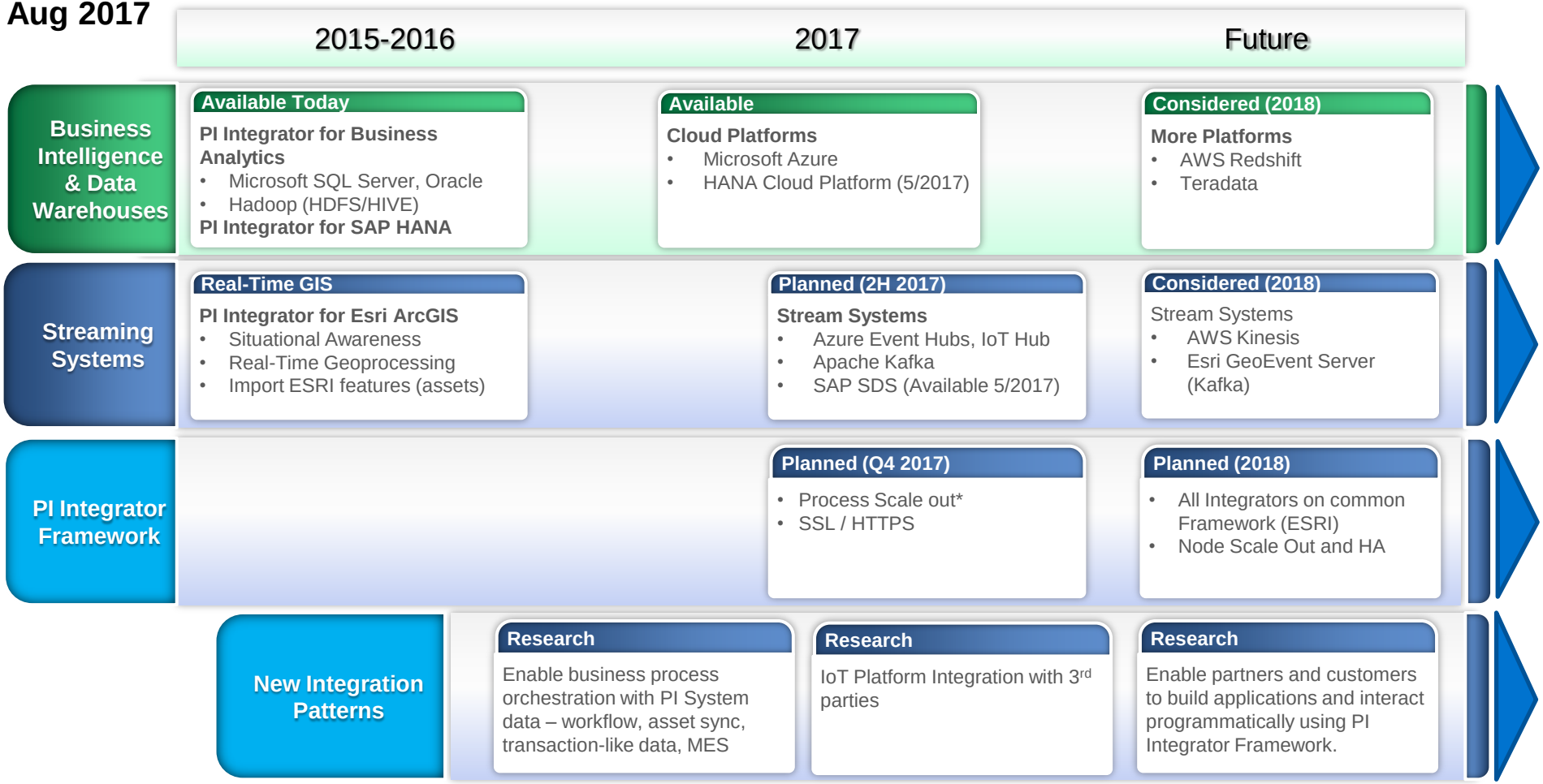
Large scale multi-variate analysis



- Machine Learning (Azure ML, R)
- PI Integrator for Business Analytics
- PI Integrator for SAP HANA

©2017 OSIsoft, LLC

Aug 2017



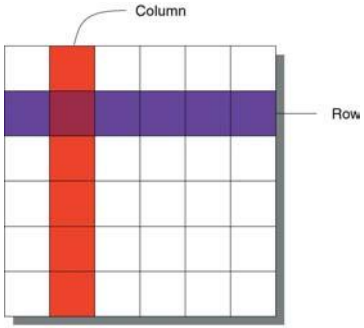
©2017 OSIsoft, LLC

PI Integrator Concepts

©2017 OSIsoft, LLC

Integration Patterns

Tables



Files
Databases

Streams

```
{
    "id":ObjectID(""),
    "name":"Phase A Angle",
    "timestamp":"2016-08-27 04:17:43",
    "value": 0.0034,
    "name":"Phase A Angle",
    "timestamp":"2016-08-27 04:17:43",
    "value": 119.9563,
    "name":"Phase A Voltage",
    "timestamp":"2016-08-27 04:17:43",
    "value": 240.0034
}
```

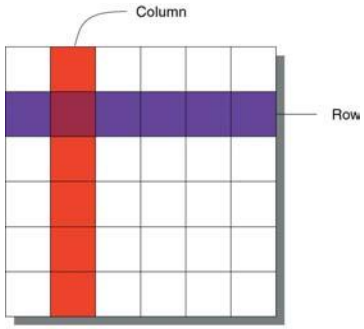
Files
Queues
Messaging
External Analytics Engines

Other Patterns

Metadata
Programming
On-Demand
Workflow & Transactions

©2017 OSIsoft, LLC

Tables vs Streams



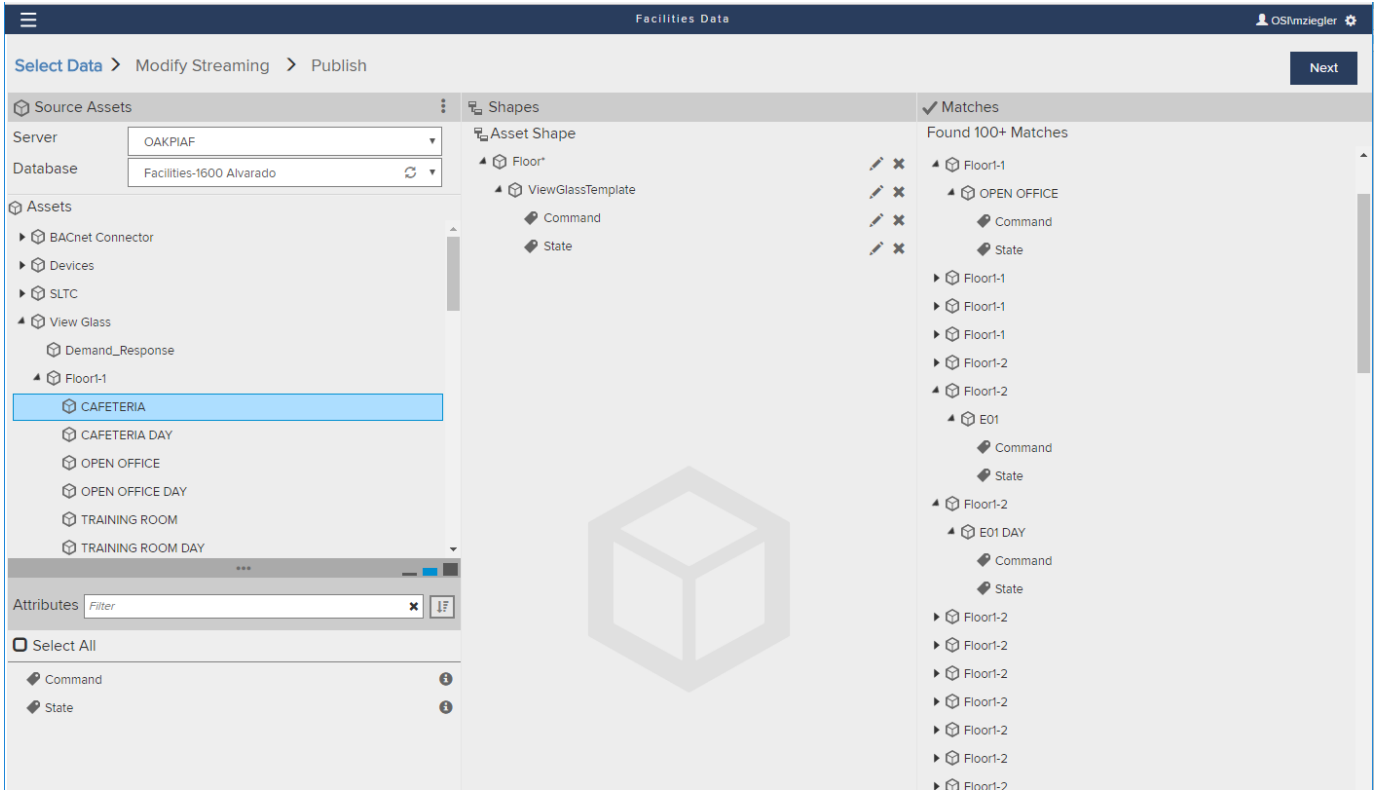

```
{
    "id":ObjectID(""),
    "name":"Phase A Angle",
    "timestamp":"2016-08-27 04:17:43",
    "value": 0.0034,
    "name":"Phase A Angle",
    "timestamp":"2016-08-27 04:17:43",
    "value": 119.9563,
    "name":"Phase A Voltage",
    "timestamp":"2016-08-27 04:17:43",
    "value": 240.0034
}
```

- Business Intelligence
- Human readable
- Batch / Bulk Process
- Normalized data
- Regularly scheduled
- Large data, few messages

- In-line (Streaming) Analytics
- Computer readable
- Specific Data / Targeted Process
- Raw or “Packages” of data
- Triggered
- Small data, many messages

©2017 OSIsoft, LLC

Data Modeling

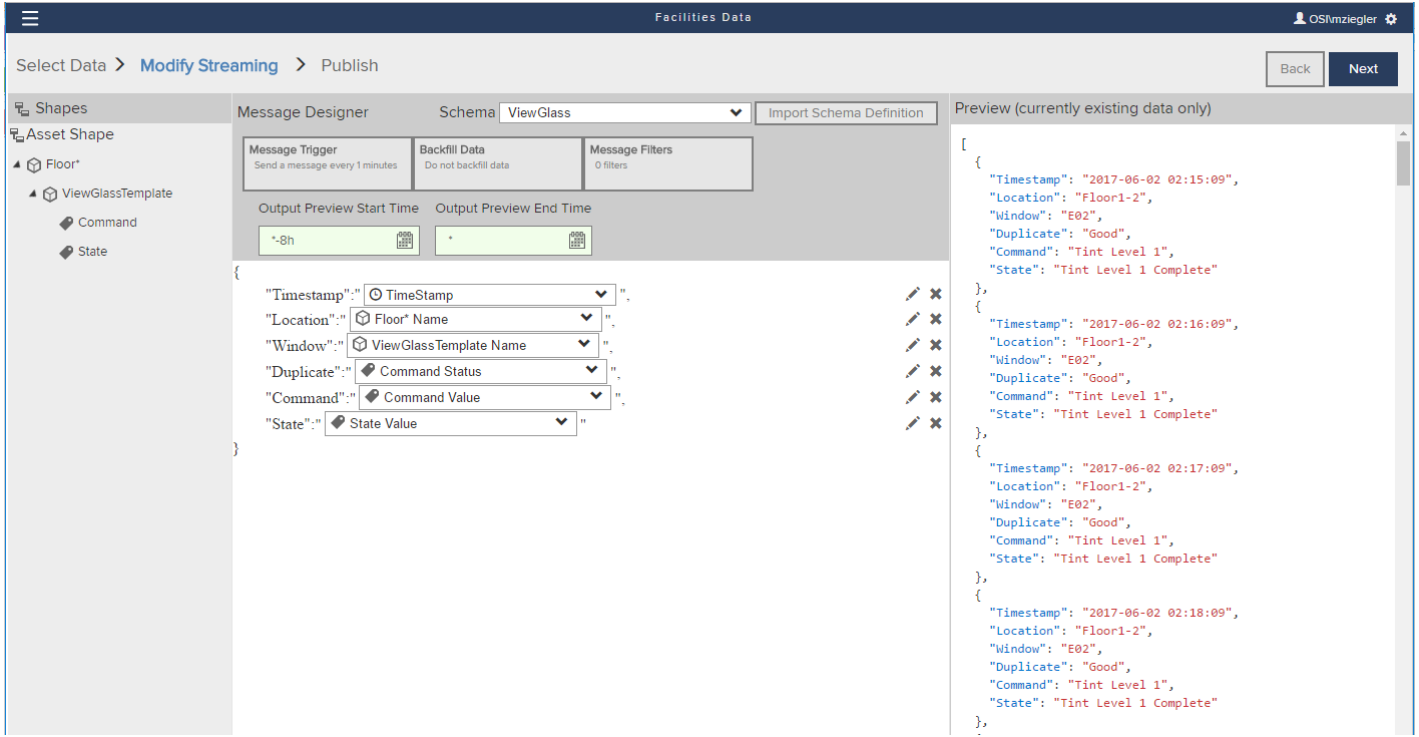


Data Modeling

- Reference the entire scope of the PI System using PI Asset Framework
- Normalize data across similar and dissimilar assets and equipment using templates, categories, and pattern searches
- Enrich datasets with location, status, calculations, and other metadata

©2017 OSIsoft, LLC

Message Designer



Message Designer

- Variety of message output formats such as JSON
- Customize the output schema, message triggering, and message content
- Filter erroneous data
- Replay historical data

©2017 OSIsoft, LLC

Use Cases

©2017 OSIsoft, LLC

Soft Measurements – Ideal Use Cases for Streaming

- Predicting Soft Measurements
  - Deschutes Brewery – Fermentation Transition
  - [Pharma Company] – Lights Out Manufacturing
  - Quality

- Enabling Data Lakes
  - Customize how data is stored & processed

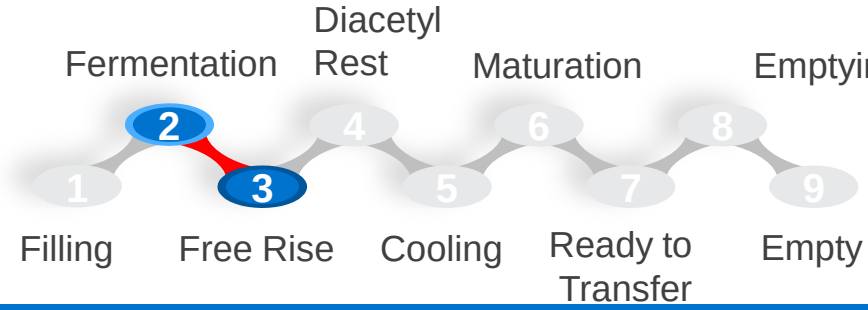
©2017 OSIsoft, LLC

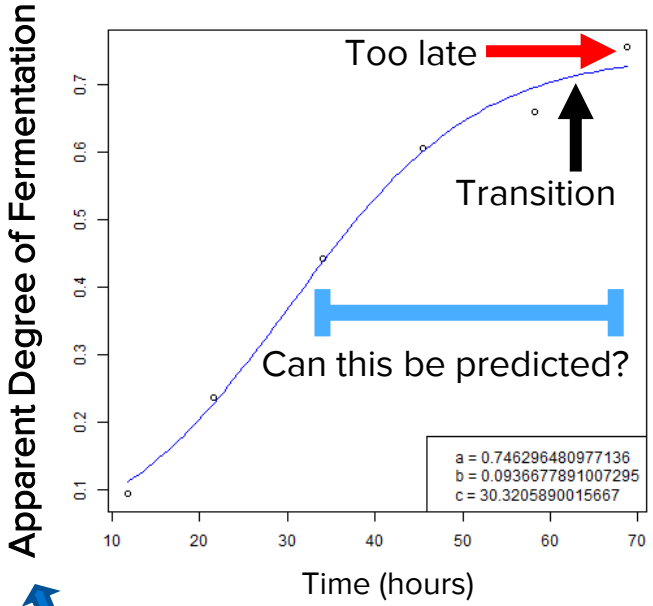
Transition occurs between infrequent manual measurements

Constraints
- One manual density measurement per vessel every 8-10 hours
- Large capital expenditure not an option

Impact
- Up to 72 hours lost in production

Options
- $750k for inline density meters
- Manually predict transition in spreadsheets

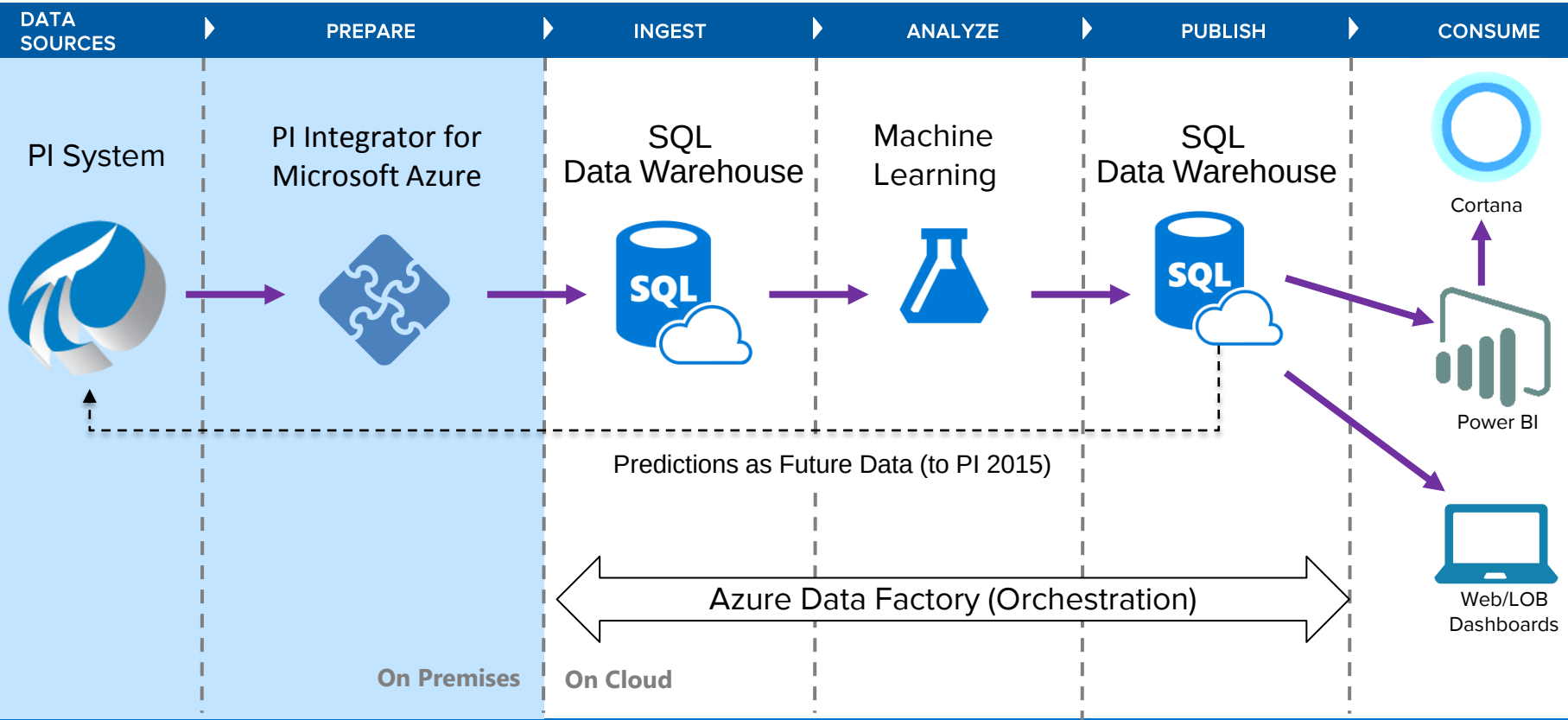




Soft Measurement

©2017 OSIsoft, LLC

Deschutes Brewery Operationalizing Architecture



©2017 OSIsoft, LLC

Automate and Optimize

Lab Measurements

- Viable Cell Count (VCC)
- Sugar Concentration
- API Concentration
- …

Online Measurements

- Agitation Speed
- pH
- **Spectroscopy**
- …



Goals: Optimize Cell Production, Eliminate Coming to the Lab

Technique: Use indirect measurements and machine learning to predict outcomes of lab measurements

©2017 OSIsoft, LLC

What success looks like for our customers

Customer	What was done	Why it matters	PI Integrator Type	Industry category
Marathon Oil	Centralized, trusted source for IT and OT data sets	• Decreased time to analyze holistic data set from **3 months to 3 weeks** • Decreased unplanned well downtime	SAP HANA	Oil & Gas
White House Utility District	Sped up identification of water leaks across a large rural area	• Optimized services team's workflow to save $30,000 per year • Saved **$900,000 in 2 years** by preventing ongoing water leaks	Esri ArcGIS	Water
CEMEX	Democratizing data for decision makers	• Reduced time to begin analysis from **months to minutes** • Reduced product variations significantly	Business Analytics	Metals, Mining, & Materials
Deschutes	Predicting events in brewing process across all brands	• **Avoided $750,000 cost** to automate density measurements • Saving on average 48 hours of production time per batch of beer	Microsoft Azure	Food & Bev
Archer Daniels Midland	Prescribe operating parameters that extend equipment's life	• Losing $50,000 every time machine fails, and 28 machines failing at a high and unexpected rate • Need advanced analytics approach to analyze 100s of variables that could be influencing failure	Business Analytics	Food & Bev

©2017 OSIsoft, LLC

Announcing Community Technology Preview

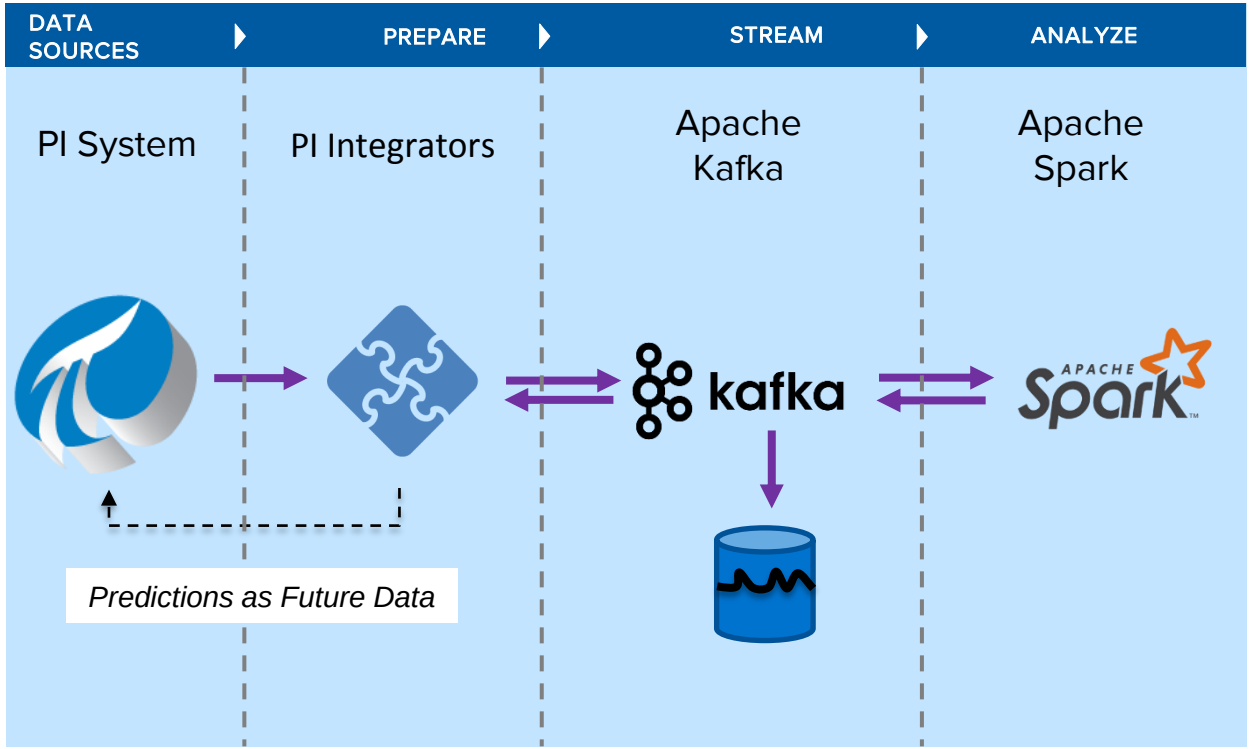
- PI Integrator for Business Analytics 2017 CTP (Open)
- PI Integrator for Microsoft Azure 2017 CTP (RCIP)

©2017 OSIsoft, LLC

Demo

©2017 OSIsoft, LLC

Reference Architecture for Streaming Analytics



Open Source Integration Pattern (Lambda Architecture)

Primary Use Cases –

- Deliver modeled input data to CEP calculations and write predictions to operational system of record
- Fill a data like and maintain 1:1 fidelity of raw data or projected data from operational system of record
- Deliver aligned array data at scheduled intervals to R or Spark based compute engines

©2017 OSIsoft, LLC

Questions

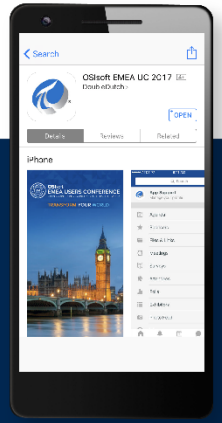
Please wait for the **microphone** before asking your questions



State your **name & company**

Please remember to…

Complete the Online Survey for this session



Download the Conference App

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

Search **OSIsoft** in the app store

Download on the App Store

GET IT ON Google Play

HTML

©2017 OSIsoft, LLC

Matt Ziegler
mziegler@osisoft.com
Product Manager
OSIsoft



Daniele Farris
dfarris@osisoft.com
System Engineer
OSIsoft



©2017 OSIsoft, LLC

감사합니다

Danke

谢谢

Merci

Gracias

Thank You

ありがとう

Спасибо

Obrigado

©2017 OSIsoft, LLC

Presented by

©2017 OSIsoft, LLC