



**OSI**soft

# **USERS CONFERENCE 2015**





# Analytics and Big Data with the PI System – Part 2: Statistical Analytics

Presented by Matt Ziegler, Product Manager, OSIsoft  
Wes Dyk, Data Scientist, Noble Energy

# PI Integrators

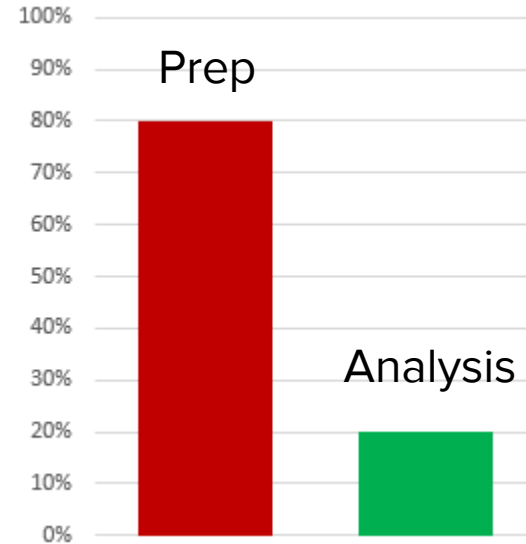
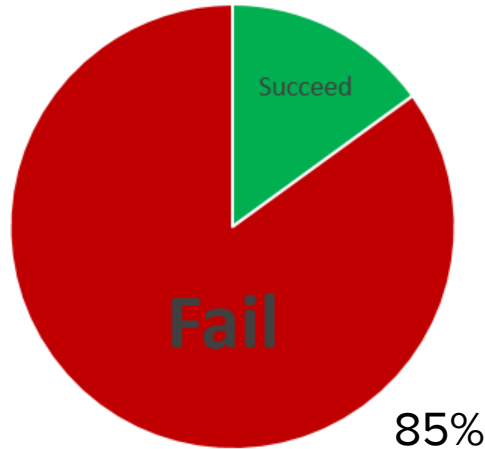
PI Integrators reduce the complexity of analyzing real world industrial data

*All PI System data delivered on **your terms**, in **your language**, to the **tools you use**, and to the people that can make a difference.*



# Big Expectations

64% of large enterprises plan to implement a big data project. 85% will be unsuccessful.



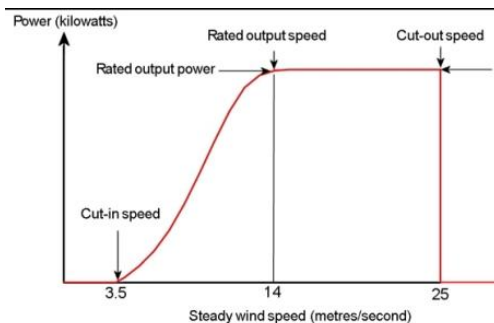
Data cleansing and preparation tasks can take 50-80% of the development time and cost.

<https://hbr.org/2014/04/the-sexiest-job-of-the-21st-century-is-tedious-and-that-needs-to-change/>





Downtime events



Startup Phases



Fleet Assets  
and Different  
Brands



Communications  
and incomplete data



Plant turn-around



# New Persona: Information Architect



- Works in an IT or OPs IT role
- Responsible for integrating many data systems to achieve business goals
- Data Lifecycle Management
  - Access Policies & Distribution
  - Logical and Physical Architecture
  - Creation and Receipt (Trust)
  - Maintenance & Disposition
- PI System is just 1 system
- Business Expert / Business Enabler / Not subject matter expert (engineer)



# New Persona: PI Data Scientist



## MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

## PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

## COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

source: <http://www.marketingdistillery.com/wp-content/uploads/2014/08/mds.png>

# Noble Energy – Company Highlights



## UNIQUE ASSETS

(Year-end 2013)

DJ Basin  
Marcellus Shale  
Deepwater Gulf of Mexico  
West Africa  
Israel and Cyprus  
Falkland Islands  
Northeast Nevada  
Levant Basin

## UNIQUE STRATEGY

Noble Energy's history began in 1932. Since then, the company has succeeded where others would not venture – applying global experience to safely and responsibly create new opportunities. We have proven our ability to move from discovery to efficient execution of large-scale development projects, and the next wave continues with six new major projects sanctioned in 2013.

Today, Noble Energy is a leading independent energy company. An S&P 500 company, Noble Energy employs more than 2,300 people and has a broad asset base that includes development and exploration opportunities in five core areas.

## NOBLE ENERGY AT A GLANCE

- Year-end 2013 proved reserves: 1.4 BBoe
- Year-end 2013 market capitalization: \$23 billion
- Projected 2014 capital program: \$4.8 billion
- Projected 2014 production: 302 - 322 MBoe/d
- NYSE: NBL

## GROWTH OUTLOOK

- Production: compound annual growth rate (CAGR) of 18% through 2018
- Proved reserves: 2.9 BBoe in 2018
- Cash flow: 19% CAGR through 2018



# Me @ Production Analysis and Optimization

## Demonstrate the potential value of Data Science

I have been charged with showing value using data science.

There is a plethora of work to build on.

We have worked to pick use cases and some just fall out of the sky.

# Me @ Production Analysis and Optimization

## I didn't start as a data scientist

I've been a developer, database administrator, system analyst, system architect, etc.

I really love math. I have a BS in Applied Mathematics and am completing a MS in the same.

Getting into the data science mindset has taken work.



# Me @ Production Analysis and Optimization

## Interdependence / Community

It is pretty straightforward to do advanced analytics, once the data is in hand.

Getting the data in the right place is difficult.

A team dedicated to all aspects of data science is essential for success.

Resources have been critical

- OSIsoft
- Hortonworks
- Academia (UC Denver and Carnegie Mellon University)



# Our Approach

## Data as a Strategic Asset

### Gain New Insights Into Production Dynamics

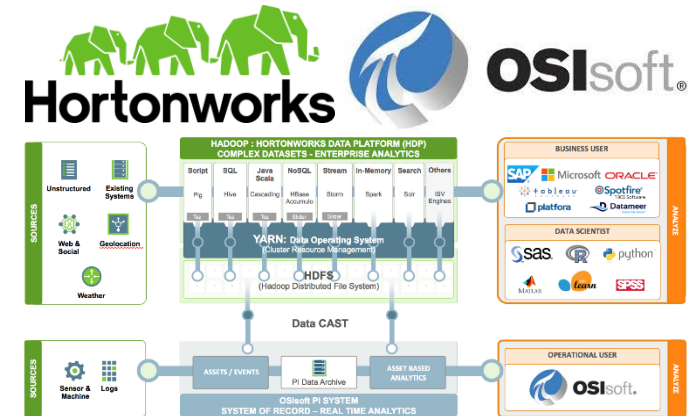
- Across a wide variety of disparate data sources and variables such as well information systems, SCADA data, sensors, and unstructured data

### Build Executive Consensus and Business Sponsorship

- Start with a single business unit – prove the value
- Recognition that data provides sustainable competitive advantage
- Widespread, systemic value creation because data is managed as professionally as capital or labor

### Rely on Trusted Partners to Assist

- Combination of Noble team members along with Hortonworks (Hadoop) and OSIsoft (PI / CAST)
- High performing operational and data science team (Python)



# Data Science / Operations Research

## Concept of a Record

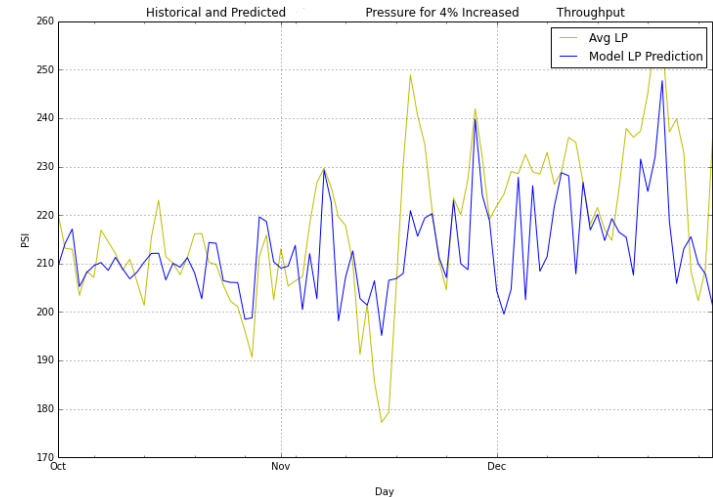
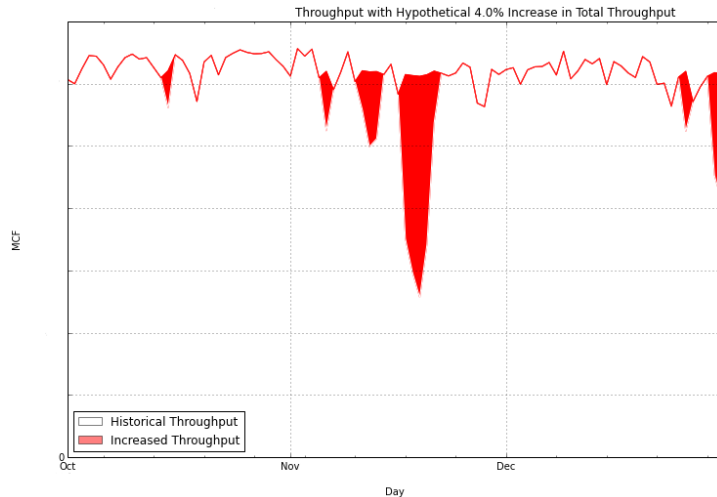
- The record or atom is essential and not much different than a row in a spreadsheet or record in a database
- Create an atom for method input
- Supervised learning
  - Collection of these records for training
  - Predictions can be made from individual atoms
- Unsupervised learning
  - Clustering of individual atoms
  - Neural Networks

# Data Science / Operations Research

## Some Exciting (for me) Projects

### Facility Downtime Analysis

- Search data for downtime (Binary Search)
- Use generated data to predict system pressure (Random Forest)
- Use system pressure to infer production volume (Direct Calculation)



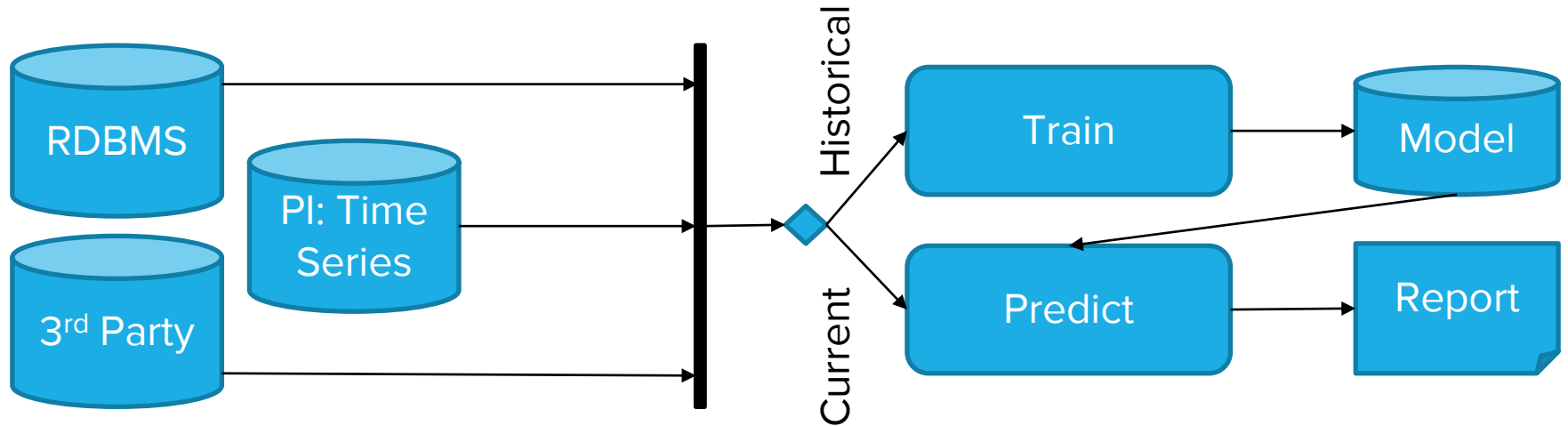


# Data Science / Operations Research

## Some Exciting (for me) Projects

### Well Downtime Analysis

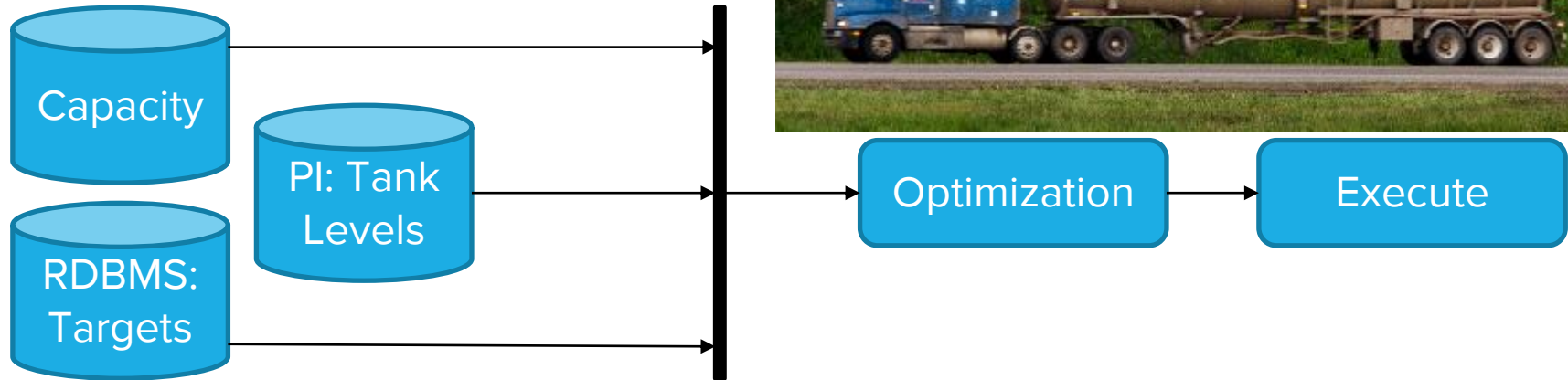
- Combine data into records for a model
- Build the predictive model from history (Random Forest)
- Make predictions using the most current data



# Data Science / Operations Research

## Some Exciting (for me) Projects Hauling and Inventory Optimization

- Mixed integer linear programming
- Decision problem using Model Predictive Control (MPC)



# Data Science Tools

## Primary Tools

- Hadoop – Hive vs Hbase vs plain HDFS
- Python vs Java

## Specialized Packages in Python

- scikit-learn for machine learning
- cvxopt for convex optimization
- GLPK for integer programming

## Challenges

- Distribution of model training
- Tools and data have been difficult to integrate

## Successes

- Ad hoc analysis
- PI Integrator has saved the day with delivering usable data sets

# Demo



## My Views

Logged in as NT AUTHORITY\USER | Logout ?

**+ Create New Asset View**  
Descriptive text can go here and span two lines.

**+ Create New Event View**  
Descriptive text can go here and span two lines.

**+ Modify View**  
Descriptive text can go here and span two lines.

**+ Remove View**  
Descriptive text can go here and span two lines.

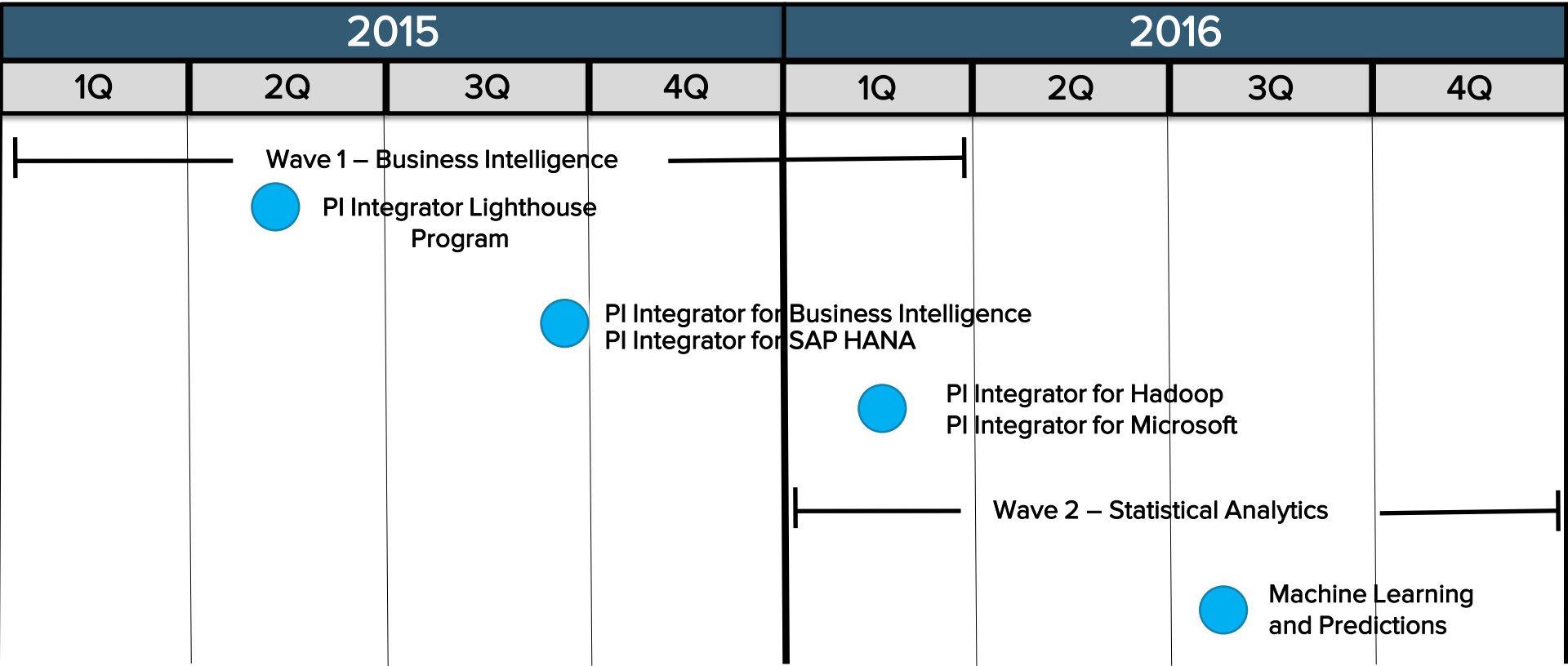
Name ▾	Last Status ▾	Type ▾	Run Mode ▾	Start Time ▾	End Time ▾	Last Run Time ▾	Rows Written ▾	≡
Daily Production		asset		0001-01-01T00:00:00	0001-01-01T00:00:00	0001-01-01T00:00:00	0	
Anacortes Heat Exc...		asset		2015-01-01T08:00:0...	2015-04-01T07:00:0...	0001-01-01T00:00:00	0	
Production Forecast		asset		0001-01-01T00:00:00	0001-01-01T00:00:00	0001-01-01T00:00:00	0	
Refinery Comparison		asset		0001-01-01T00:00:00	0001-01-01T00:00:00	0001-01-01T00:00:00	0	
Heat Exchanger Re...		asset		0001-01-01T00:00:00	0001-01-01T00:00:00	0001-01-01T00:00:00	0	
Pumps		asset		0001-01-01T00:00:00	0001-01-01T00:00:00	0001-01-01T00:00:00	0	

[Overview](#) [Data](#) [Config](#) [Performance](#) [Log](#) [Security](#)

Time ▾	View ▾	Level ▾	Message ▾	≡
2015-04-23T15:43:31.52366...	TestThursday	Informational	AFReader[0]:Init: Initialization Completed Successfully.	
2015-04-23T15:43:31.651731	TestThursday	Informational	AFReader[0]:Open: AF Server='X-AF' - Connected.	
2015-04-23T15:43:31.65652...	TestThursday	Informational	AFReader[0]:Open: AF Database='ContinuousTest' - Found.	
2015-04-23T15:43:31.7412249	TestThursday	Informational	ShapeInstance: EntryPoint (Auto) set to: [1]	

# PI Integrators Roadmap

*Subject to change*





# V1.0 Features – Q3 2015

## Data perfectly suited for BI

- **Row Column Data (Dimensional model)**
  - Data alignments: Evenly spaced in Time, Reference attribute Time, By Events
- **Comprehensive coverage of AF**
- **Event Frames – As Filters, As Data**

## View updates on a schedule

- **Caveats: V1.0 does not detect history changes**

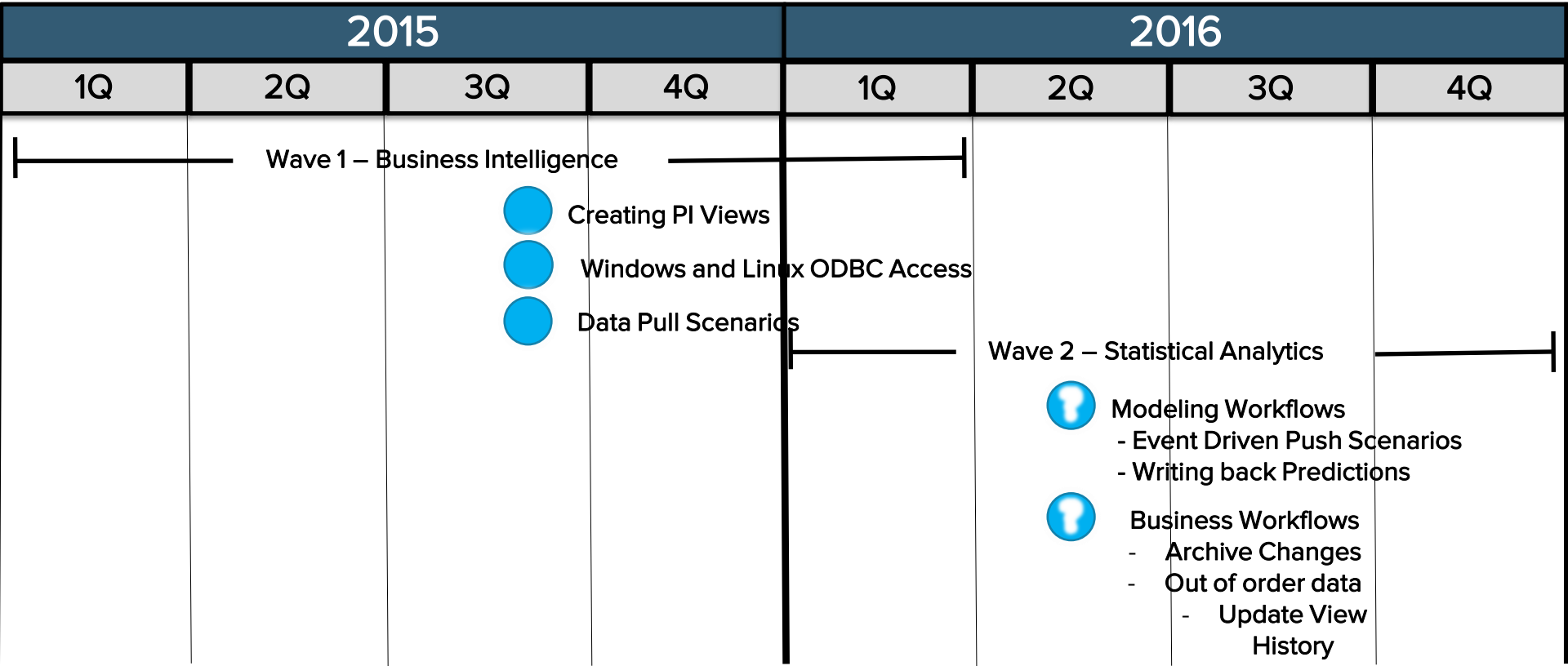
# V1.0 Features – Q3 2015

## Platform Integration

- **SAP HANA**
  - Automatically creates Virtual Tables via Smart Data Access (SDA)
- **PI View**
  - Access decision ready data via ODBC with first class support for Tableau, Spotfire, and Microsoft BI (Excel, PowerPivot, PowerView, Power BI)
  - Works with any software with an ODBC data source connection

# Feature Focus Areas

*Subject to change*



# PI Integrator Lighthouse Program

Jump start your big data or analytics project with help from OSIsoft. [PIIntegrators@osisoft.com](mailto:PIIntegrators@osisoft.com)

<https://www.surveymonkey.com/s/PXW5L6X>

6 week engagements through September 2015.

# Questions

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

State your  
**name & company**





THANK  
YOU





# Matt Ziegler

[mziegler@osisoft.com](mailto:mziegler@osisoft.com)

Product Manager

OSIsoft, LLC



# Wes Dyk

[Wes.Dyk@nblenergy.com](mailto:Wes.Dyk@nblenergy.com)

Noble Energy





**OSI**soft

# **USERS CONFERENCE 2015**

