

Bringing Big Data To New Verticals

Jerry Schuman
Chief Technology Officer


Sean Patrick Murphy
Chief Data Scientist

October 23, 2014

“Most of us are trained to believe theory must originate in the human mind based on prior theory, with data then gathered to demonstrate the validity of the theory. Machine learning turns this process around. Given a large trove of data, the computer taunts us by saying, If only you knew what question to ask me, I would give you some very interesting answers based on the data. Such a capability is powerful since we often do not know what question to ask.”

— Dr. Vasant Dhar

*Center for Data Science at New York University
Stern School of Business*



Our parent fund Frost Data Capital
brings industrial point solutions
that are anchored in big data and
machine learning.

Oil and Gas

Oil & Gas Case Study

INDUSTRY

Sensor Data Analytics

SUMMARY

Bringing together real-time big-data technology and innovation to machine-learning for physical systems to transform the management of high-value industrial assets

Asset-based Analytics

Upstream Production, Drilling and Completion

Recognizes Operation and Failure Modes for ESPs, Rod Pumps, Compressors and Frack Jobs

Customer focus on Reducing NPV by 30%

Tackles a \$104 billion problem



New Opportunity in Oil & Gas

INDUSTRY

Oil & Gas Upstream

SUMMARY

Optimizing a complex, multi-stage well delivery process across disparate information silos

Adaptive Well Delivery

Siting, drilling, fracking and completing a well demands coordinated operations across many different functional groups

Each group is optimized individually, but little coordination to find the best overall process

E.g. a small change in drilling might cost more, but have an enormous impact on total well production

Solution: use domain-specific languages to capture complex interdependencies between groups and find a global optimum



Mining



USE CASE #1:

Grinding Circuit Optimization – Throughput and Electricity Usage



Grinding Represents 40% of Energy Used by Equipment

Holistic Optimization:
Minimize Bottlenecks &
Maximize Throughput

Better Forecasting & Reduce
Spinning Excess

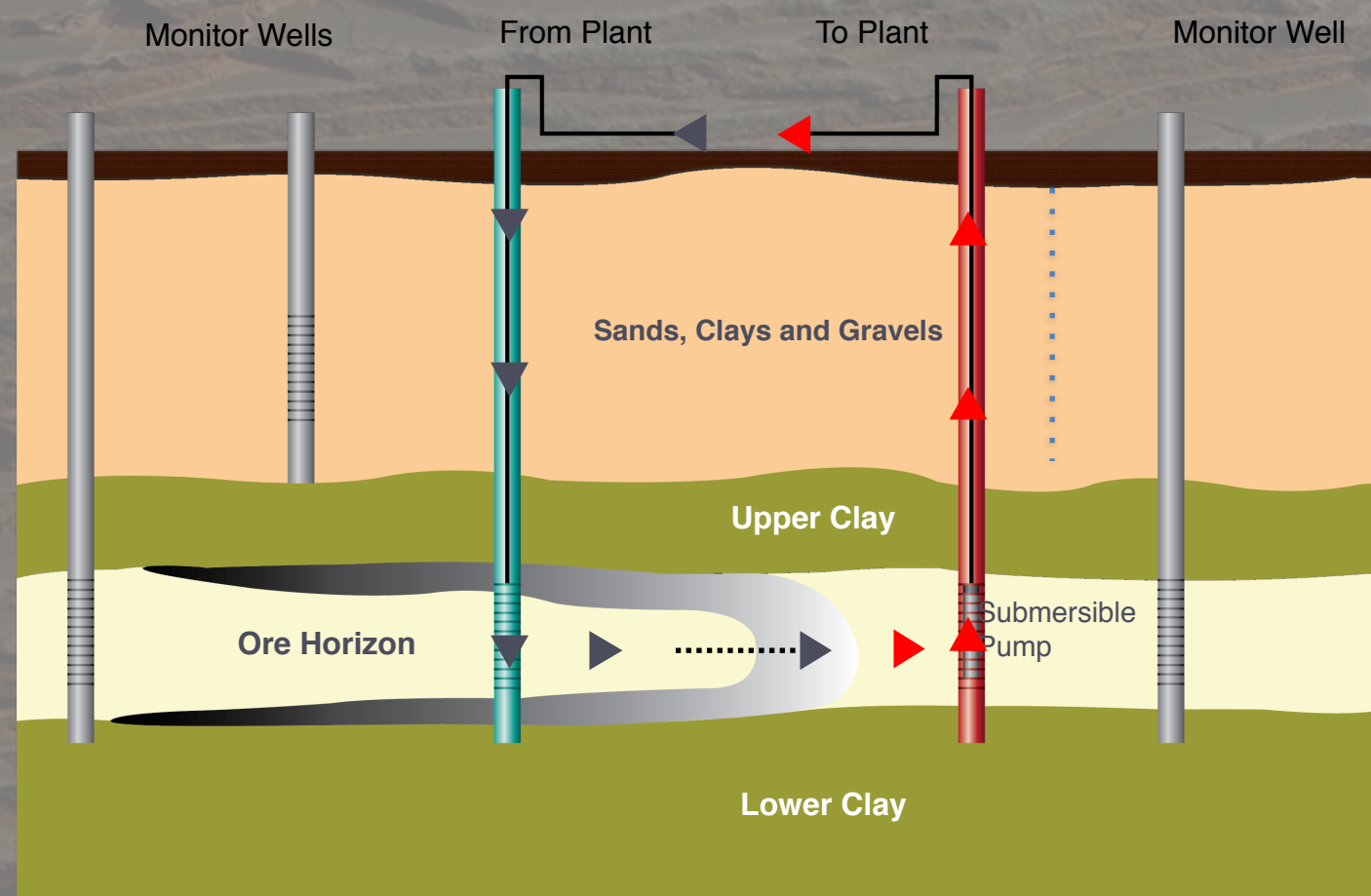
Potential Savings: 667 Trillion BTU Annually Across Industry





USE CASE #1:

In-Situ Uranium, Recovery Optimization & Reduction of Downtime



Optimize Pump Efficiency & Production

Optimize Lixiviant & Water Utilization


Predict/Prevent
Depressurization From
Aquifer Breaches

10-20% Productivity Gains = \$150M - \$300M Annually in Uranium



HealthCare



 JointlyHealth

David Director [Settings](#) [? Help](#) [Logout](#)

[←](#) Wellness Model [Edit Properties](#)

Clinical Rules **Compliance Rules**

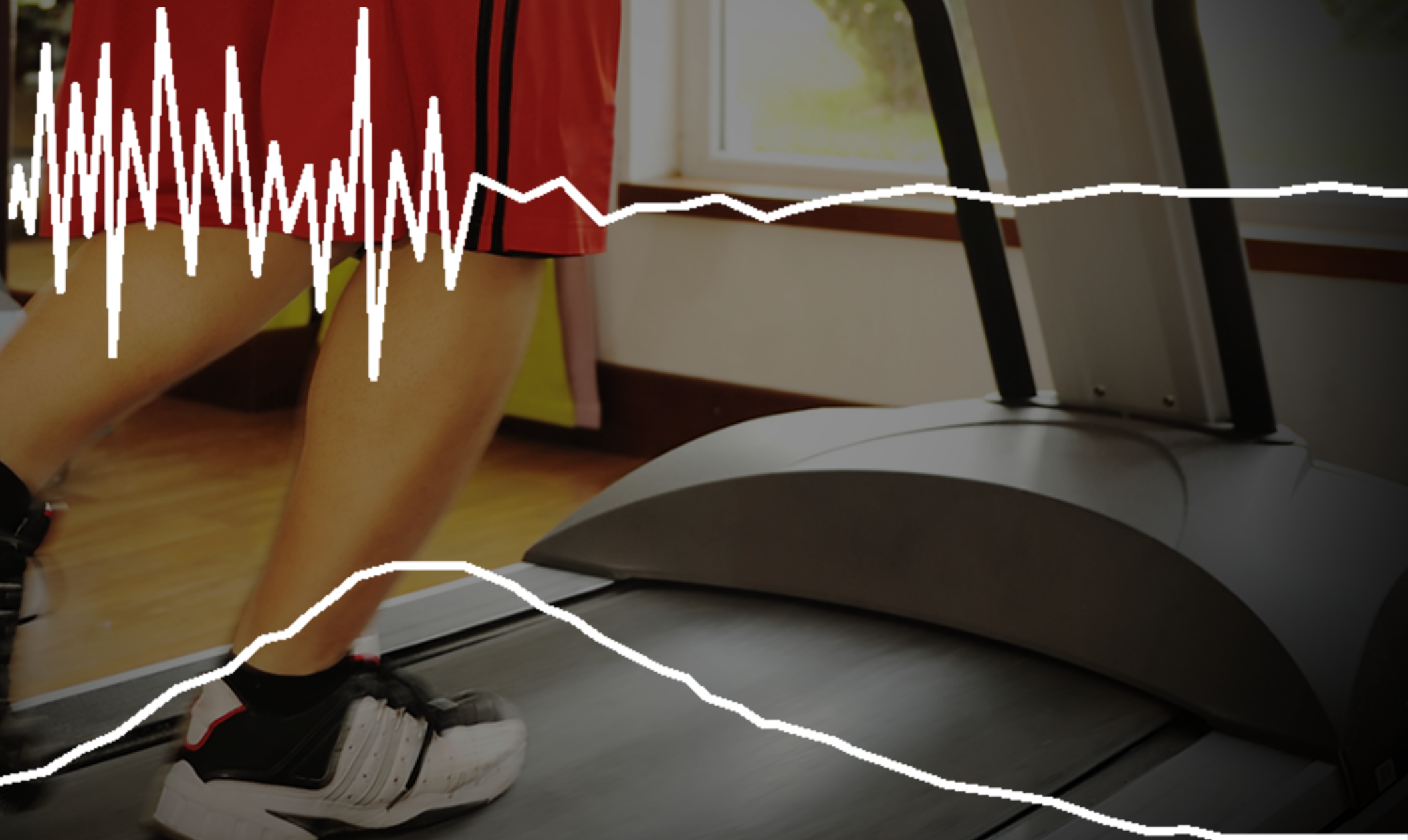
[Add Rule](#)

Rule Name	Expression	Risk Category	Edit	Delete
<input type="checkbox"/> High Blood Pressure	The last measurement of morning Blood Pressure (Systolic) is above the value of 170 mmHg	<div></div>	✎	✖
<input type="checkbox"/> Weight Gain	The 4 day average morning Weight is above the 4 week baseline by 5 %	<div></div>	✎	✖
<input type="checkbox"/> Resting Pulse	The 3 day average Pulse Rate (Vital Connect Patch) is above the 4 week baseline by 2 standard deviations	<div></div>	✎	✖

[Split](#) [Combine](#)

Activity Level

Heart Rate



Activity Level

Heart Rate

Machine Learning





Timeline 24 Hours 3 Days 5 Days 15 Days 1 Month 6 Months

Patient Since Dec 16, 2013

Last 30 Days 2 Alerts

Events

EVENT DETAILS

[illegible]

1:01 AM SpO2 = 85%

Graphs

Tables

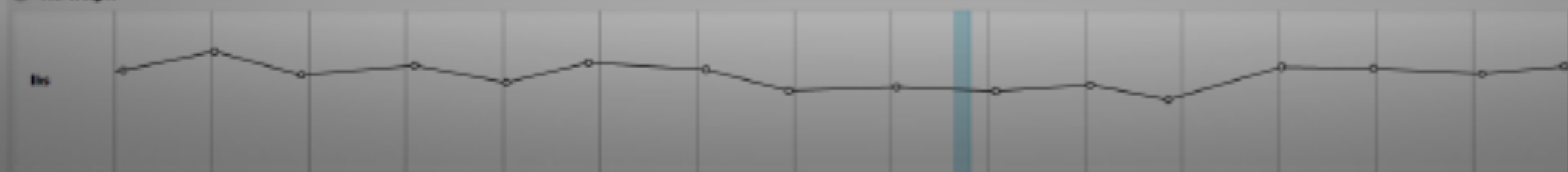
● Post-Exercise HR



Machine learning Alert:

■ HR recovery pattern:
Max HR provided 280 bpm
Recovery pattern deviation

AM Weight



No Contribution

AM Thoracic Fluid

AM Symbolic BP


Key Contributions

Review Clinical Event

Event HR recovery differs from baseline profile.

Patient , Sam

Sent Jul 25, 2014, 4:17 AM

Risk Level 

Current State unread

Related Data Readings

O2 94% (3 Day AVG)

Case Management Notes

Enter note here

Clear Note

Save Note

Added By

When ↓

Note

Event Resolution

Event Status ☐ In Progress ☐ Resolved

Where is the starting
line?

Data Providers

- SCADA
- PMU
- Specific Instrumentation
- Log Files
- Databases
- External Third Party Services
- Documentation

Data Formats

- Structured
 - Databases - pick your favorite flavor
- Unstructured
 - Documents, Reports, images...
- Semi-Structured
 - Log Files
- Time Series Data
 - SCADA, PMU etc

tools, lots of tools



Getting Started

With Big Data, the value is discovered through iteration and refinement: pose a question, create statistical, visual, or semantic models, validate, then ask a new question.

“Discovering meaning in your data is not always straightforward. Sometimes, we don’t even know what we are looking for initially. That’s completely expected. Management and IT needs to support this “lack of direction” or “lack of clear requirement.” So, to accommodate the interactive exploration of data and the experimentation of statistical algorithms we need high performance work areas.”

Boom or Bust

Did we mention that this is about exploration?

Expect the unexpected and expected

Going Forward

The Approach

Expert-Augmented Learning

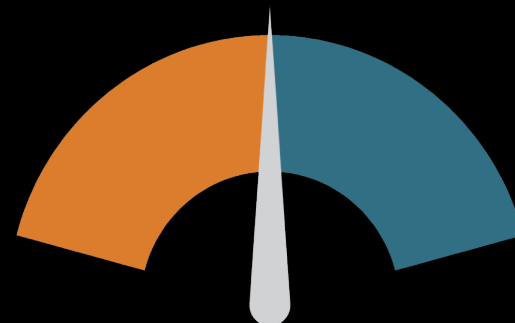
Case Based Reason

Similarity Based Modeling



HIGHLY VARIABLE
ENVIRONMENT

HIGHLY DIVERSE
ASSETS



MIXED
ENVIRONMENT

MIXASSETS

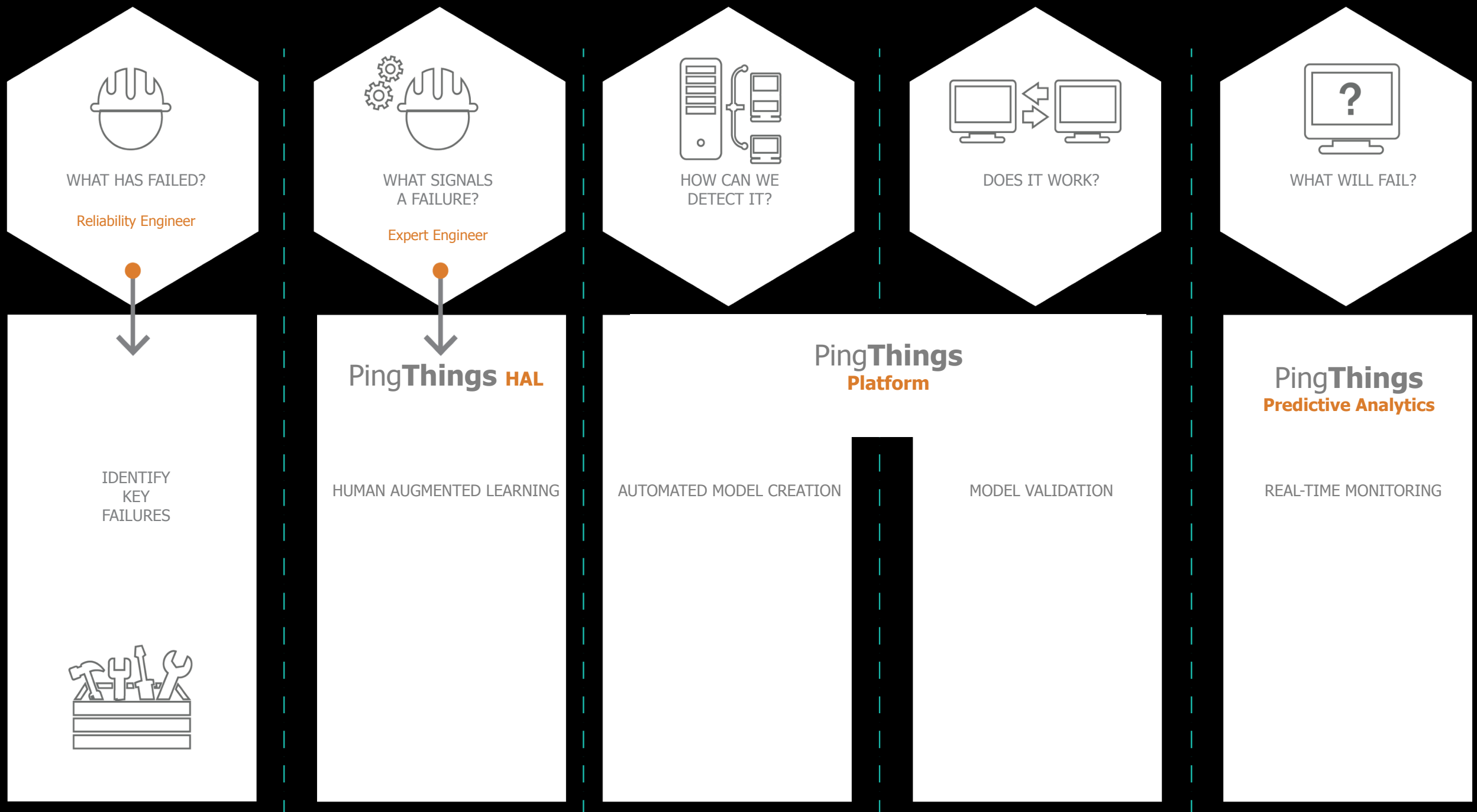


HOMOGENOUS
ENVIRONMENT

IDENTICAL
ASSETS



The Approach



Big Data and machine learning brings, for the first time, Operational Intelligence to the industrial sectors.

Now, you just need to ask questions and don't stop asking.