Recursive Splunking

Building a data Rube Goldberg machine with Splunk

Greg Hrebek

Director of Technology,
New York Air Brake

splunk>

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Agenda

- Who We Are and What We Do
- The Challenges of Riding the Rails
- Creating a Single Source of Truth
- Recursive Splunking
- Assembling the Machine
- Beyond the Obvious

Who We Are and What We Do

splunk>

Who Am I?

- Currently the Director of Engineering for New York Air Brake
- Previous companies - GE, Invensys, NASA, HP
- Advanced degrees in Computer Software and Systems Engineering
- Technical expertise in Embedded Systems, Safety Critical Systems, Data Analytics, Mobile, and Distributed Systems
- Have held positions in every aspect of the product development lifecycle

Who is New York Air Brake?

- A wholly owned subsidiary of Knorr Bremse
  - Knorr-Bremse is the world's leading manufacturer of braking systems for rail and commercial vehicles
  - Two pillars of business – rail and commercial vehicles
  - Privately held with an average of 7% of return into new product development
- New York Air Brake – serving the rail industry since 1890
  - Worldwide leader in railroad brake and train control systems
  - 800 Employees across five divisions

Better Yet, Who is Train Dynamic Systems?

- Worldwide leading industry experts in train dynamics and training simulators, a satellite division of New York Air Brake
- Responsible for R&D and NPD for New York Air Brake
- 150 Employees
- Premier product is LEADER®, an energy management control system

The Challenges of Riding the Rails

splunk>

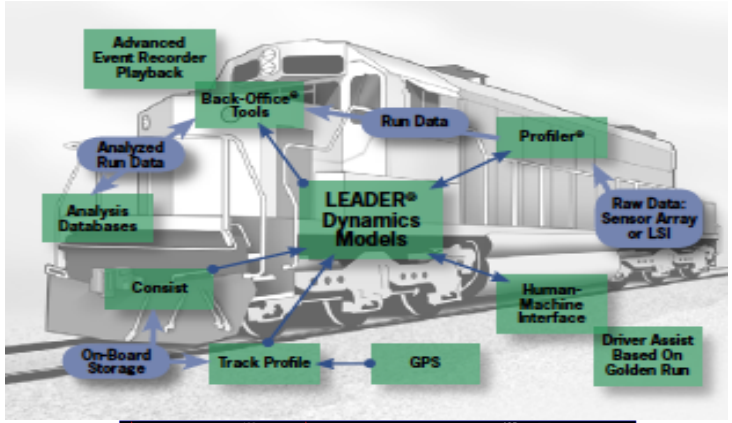
How a Train Works

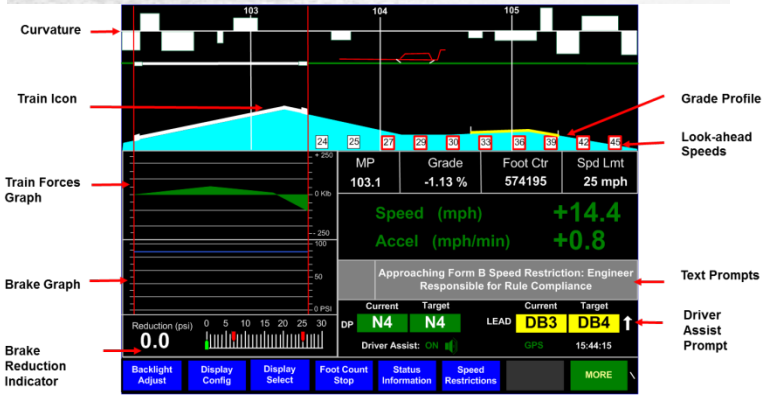
- Diesel Electric Motors
  - Diesel powered generators create electricity that is then sent to large electric traction motors. These motors, when put in reverse, will create “dynamic braking”
- Airbrakes
  - An air hose runs down the length of the train charged to 90 psi. If the air is let out, the brakes apply at each car
- In-Train Forces
  - There are about 8-12 inches of slack between each knuckle (the part that connects the train cars together), this equates to about 300-400 feet of slack in the whole train
- Controlling Momentum
  - Driving a train is like controlling a giant slinky

LEADER® Driving Strategy Engine

On board the Locomotive

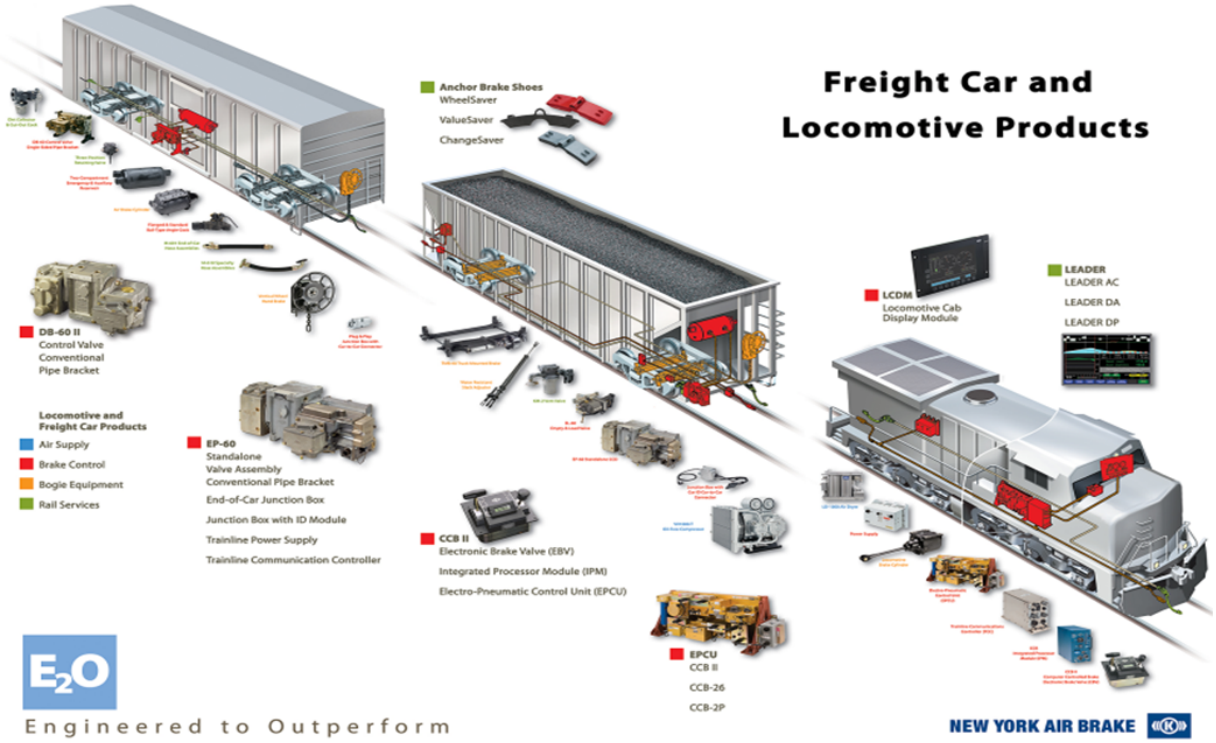
- Collects sensor data from the traction control system, brake system, and positioning system
- Provides information to the locomotive engineer that he typically would have to feel out.
- LEADER® logs all the inputs/outputs for post run analysis
- LEADER® adaptively tunes its strategy based on local conditions





Train Systems

- Locomotive Systems
  - Air Compressors
  - Traction Motors
  - Diesel Generators
  - Brake Systems
  - Computing Platforms
- Car Systems
  - Brake Shoes
  - Brake Valves
  - Bearings
  - Draft Gears



Wayside Systems

- Signaling & Crossings
  - Crossing arms
  - Rail switches
  - Weigh-in-motion scales
- Detection
  - Track circuits
  - AEI Tag readers
  - Hot/cold wheels
  - Dragging equipment
  - Track fouling
  - Wheel detectors

Back Office Systems

- Dispatching
  - Movement planning
  - Speed restrictions
  - Consist data
  - Interoperability
- Metrics
  - Network velocity
  - Trip planning
  - Asset exchanges
  - Commodity planning

The Tower of Babel

- Assets & devices range from new to 40 years old
  - Average asset life of 20+ years, leaving assets as old as 40 years in active use
- Everything speaks a different language
  - A plethora of protocols exist, few are open source or based on industrial standards, majority are proprietary and closed source
- No unifying body to drive standardization
  - Even the same manufacturer has several different protocols for active units
- No centralized aggregation
  - Assets are digitally isolated and scattered (on board, wayside, back office)
  - Large areas of migration as assets move not only within networks but across
  - Limited connectivity between assets and between assets and a back end
- Mix between machine data and calculated data
  - Data from assets and data from models & calculations

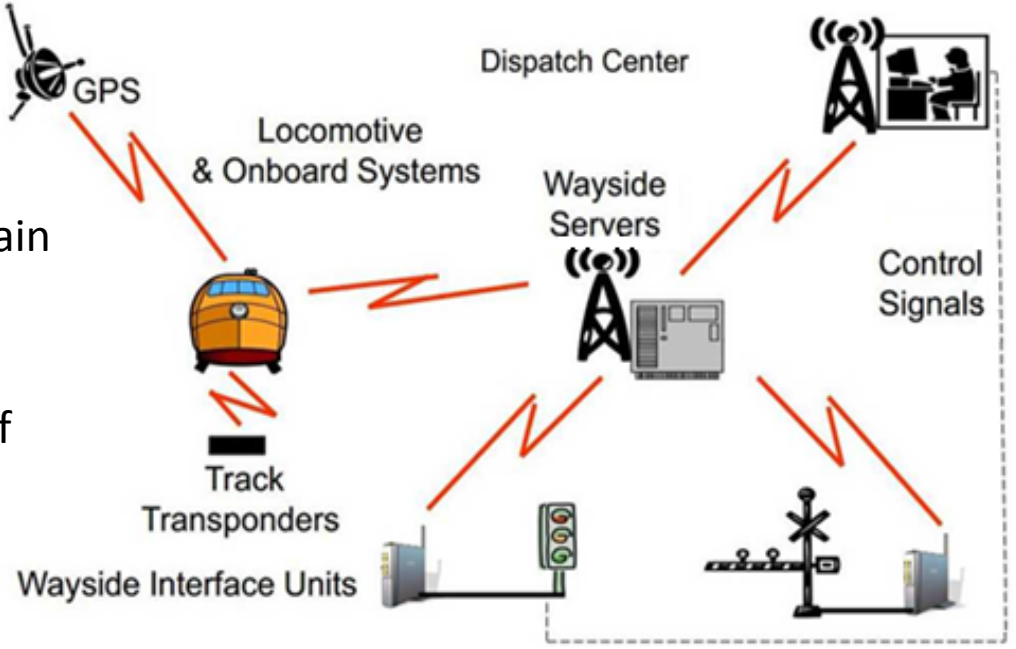
Creating a Single Source of Truth

splunk>

The Rube Goldberg Approach

- Collect and aggregate the data
- Normalize the data in the time domain
- Do something interesting with the data, aka: Feature Engineering
- Present a unified and holistic view of information and actions

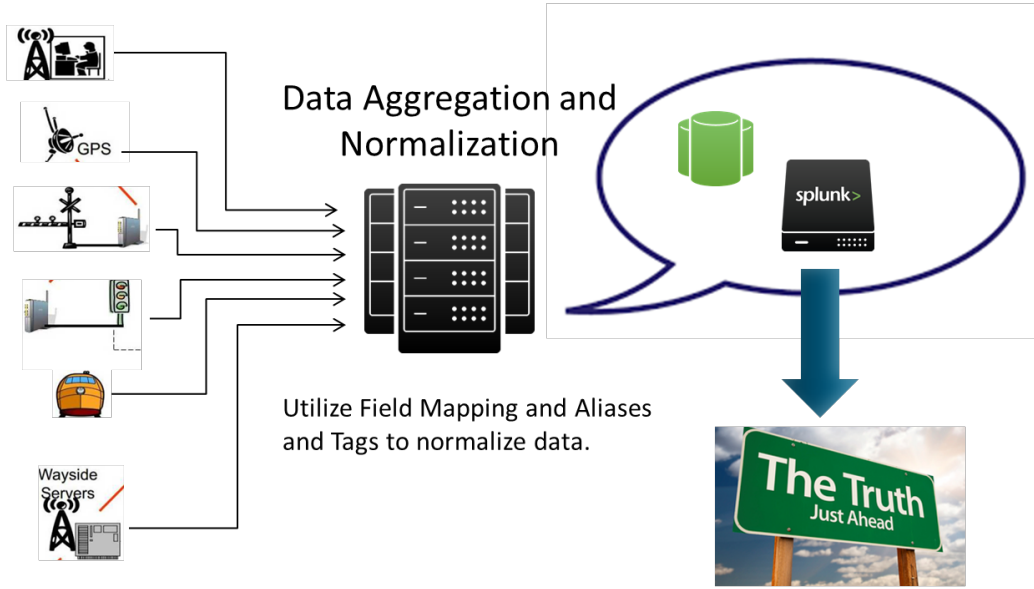
  (that hopefully people will pay for)



Aggregating & Normalizing the Data

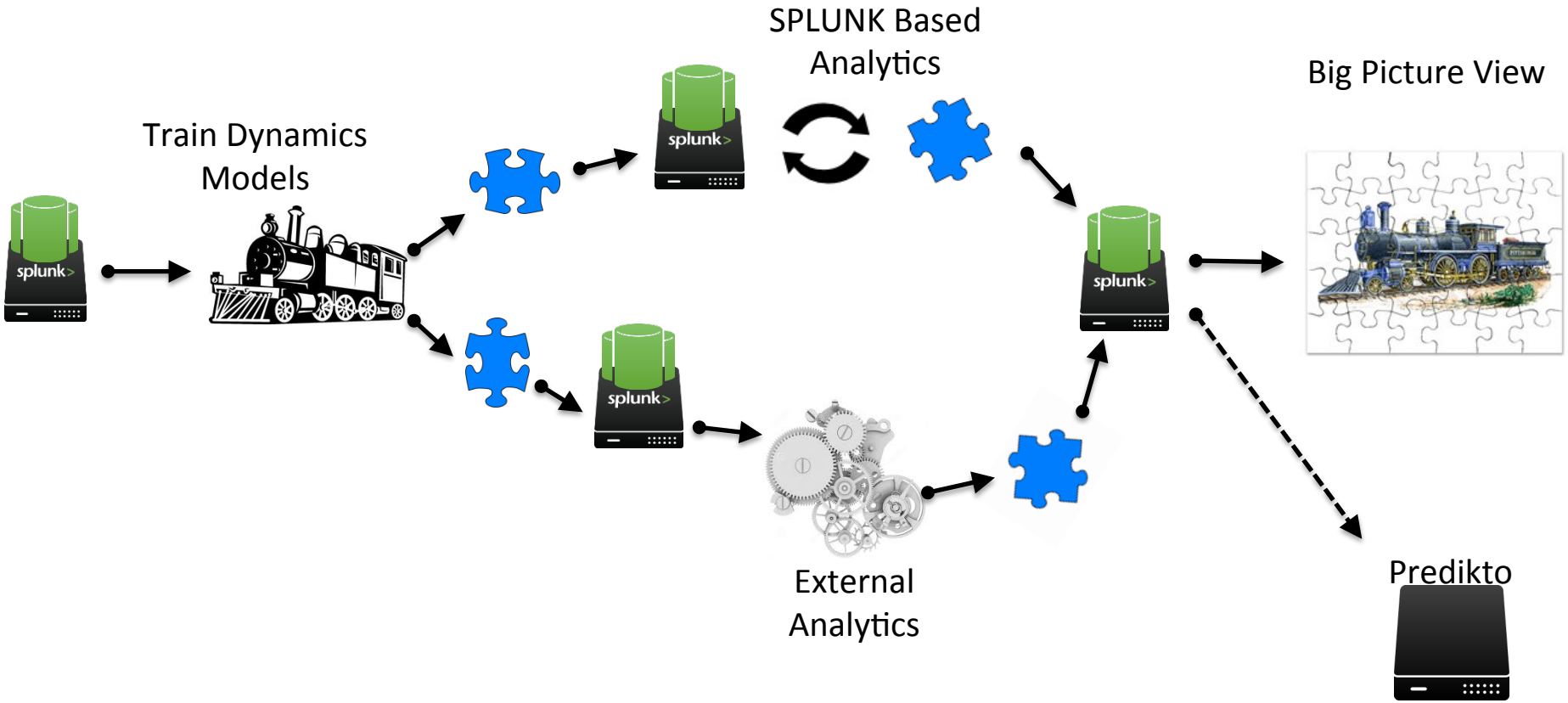
Bringing it all together: Analysis

- Aggregation schema
  - Index: customer (UP, NS, CSX, etc.)
  - Host: asset Id (Loco Id, etc.)
  - Source Type: Asset Type (ER, TPA, 10th Track, DynOut, LEADER Logs, Sensor Data, etc.)
- Field mapping & aliasing
- Data transform
  - Preprocess data inputs using dynamic data from Splunk to do so



Recursive Splunking

splunk>


SPLUNK Based
Analytics
Big Picture View
Train Dynamics
Models
splunk>
splunk>
splunk>
splunk>
External
Analytics
Predikto

Extending the Splunk Query Language

- Functions specific to our analytical needs did not exist
  - Finite impulse response filters
  - Multi-variable regressions
  - Fast fourier transforms
  - Multi-objective parametric optimization
  - Call external dynamics models
- Used Python to perform these analysis in Splunk
  - With many high level analytical libraries, Python was perfect match
  - Allowed for data management, organization, normalization in Splunk
  - Seamless interface to analytical algorithms that were already developed
  - Very well documented in docs.splunk.com

Extending the Splunk Query Language

- Intersplunk
  - http://docs.splunk.com/Documentation/Splunk/6.2.4/AdvancedDev/SearchScripts#Build_your_search_command_in_Python
  - API For accessing results from the search pipeline in Python
  - Example:

```
8  import sys
9  import splunk.Intersplunk
10
11 results = splunk.Intersplunk.readResults(None,None,True)
```

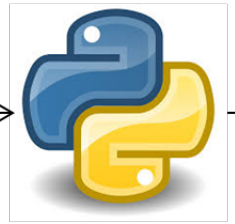
Parameter	Default	Description
inputbuf = None | file	None	Indicates where to read input from. Set to None by default, which means your search command expects to get data from sys.stdin.
settings = None | *dict*	None	Indicates where to store any information found in the input header. Set to None by default, which means do not record the settings.
has_header = True | False	True	Indicates whether or not to expect an input header.

- commands.conf
  - http://docs.splunk.com/Documentation/Splunk/6.2.4/Admin/Commandsconf
  - $SPLUNK_HOME/etc/apps/<app_name>/local/commands.conf
  - File is used to tell Splunk what scripts are available for the specific app
  - Add a reference in the file to your script →
  - Script should be in $SPLUNK_HOME/etc/searchscripts/

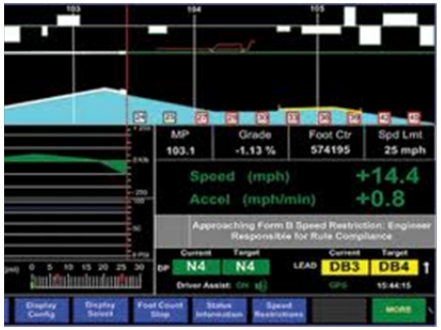
```
[diff]
FILENAME = diff.py
```

Utilizing Data Models – External Process Calls

Index = UP, Host = 8467
sourcetype = ClassD |



Intersplunk API
Python launches
QNX 6 VM
running LEADER
from search
results



LEADER recreates
train run

Intersplunk API
sends LEADER
results back to
Splunk

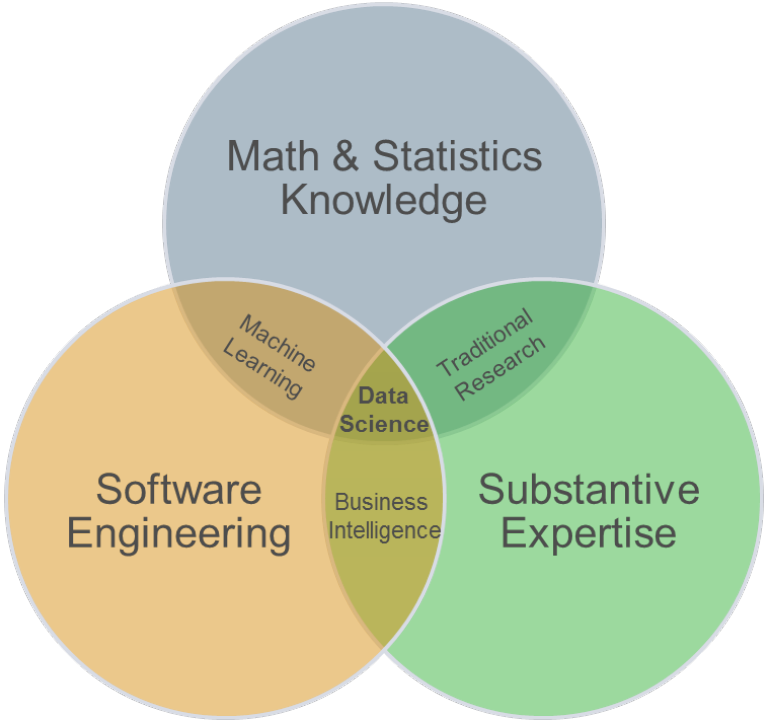
Use Python to manage data distribution wrap calls to external programs

Assembling the Machine

splunk>

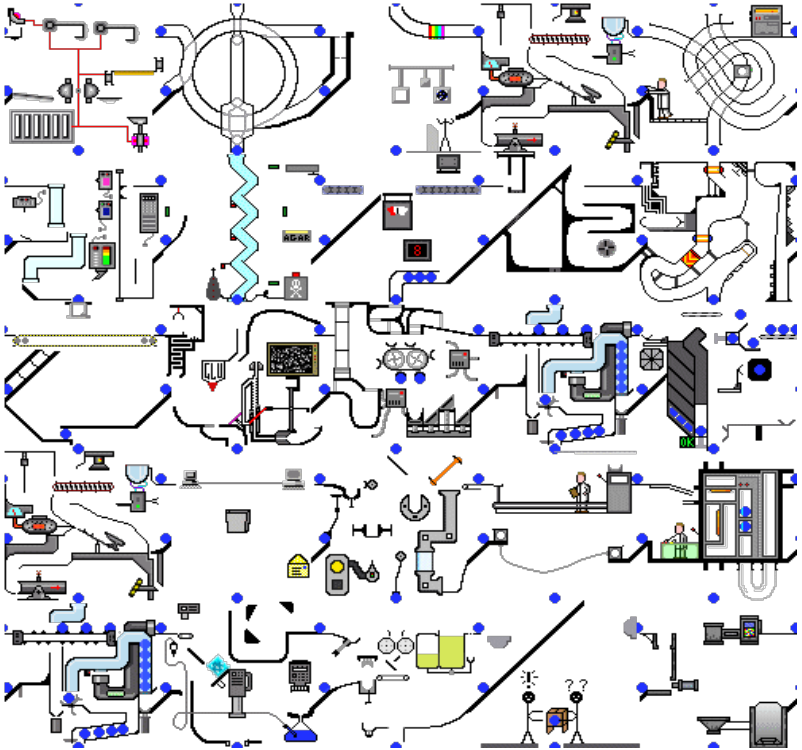
Splunk as a Data Science Platform

- Business intelligence
  - Domain expertise to answer questions
  - Particular approaches to particular problems
  - Splunk alerts and search query drill downs
- Statistics
  - The raw math and traditional analysis
  - Models and hysteresis
  - Integration by extending Splunk query language and external processes
- Data(base) management
  - Data munging and storage
  - Parsing, aggregating, and formatting data
  - Splunk field maps, aliases, scripted inputs, etc.
- Visualizations
  - Dashboards and reporting
  - BI Tool integration
- Machine Learning
  - Predictive and prescriptive analytics
  - Predikto



So Why Did We Build This?

- Fuel analysis
  - Technology Impact
  - Environmental vs. behavioral impact
  - Trending & predictive usage
- Train performance analysis
  - Tractive effort efficiency
  - Braking efficiency and effectiveness
  - Rolling resistance
  - Environment impacts (wind, temperature, etc.)
  - Un-desirable emergency opportunity warning system
- Predictive/prescriptive maintenance
  - Locomotive health
  - Track health
  - Brake shoe wear
- Metrics and reporting
  - Real-time alerts
  - Network velocity
  - Asset utilization
  - Playback
- Driving strategy tuning & analysis
  - Behavioral analysis
  - Hyperlocal optimizations



Beyond the Obvious

splunk>

SPLUNK> as a Platform

Internet of Things and More!

- It is more than just for servers and their logs, it is for building dynamic data relationships regardless of when and where the data comes from!
- Build, discover, and visualize correlations between the things that matter to your product or business
- Focus on finding answers

Want More?

- Recommended Sessions
  - Building Powerful Analytics with Ease | Tuesday, September 22, 2015 | Breakout 5
  - Deeper insights into Water Treatment Through Splunk | Tuesday, September 22, 2015 | Breakout 6
  - How Splunk Uses the Splunk Add-on for JIRA | Tuesday, September 22, 2015 | Breakout 5
  - Robotics Analytics at Target: Utilizing Machine Data from Robots to Provide Data-driven Insights and Decisions | Tuesday, September 22, 2015 | Breakout 4
  - Splunk as a Platform for Operational Intelligence for SCADA and Other Industrial Systems | Tuesday, September 22, 2015 | Breakout 1
  - Enhancing Dashboards with JavaScript! | Wednesday, September 23, 2015 | Breakout 9
  - Know Your Data, Know Your Audience | Wednesday, September 23, 2015 | Breakout 12
  - Machine Learning and Analytics in Splunk Wednesday | September 23, 2015 | Breakout 11
  - Of Babe Ruth and Spider Man: How Baseball and Comics Can Teach You the Search Language | Wednesday, September 23, 2015 | Breakout 9
  - The Internet of Things...Predicting the Unpredictable | Wednesday, September 23, 2015 | Breakout 14
  - Unraveling Analytics and Data Science: An Expert Panel | Wednesday, September 23, 2015 | Breakout 13
- Contact Info
  - Greg.Hrebek@nyab.com
  - http://www.splunk.com/view/SP-CAAAKZS
  - http://www.splunk.com/view/splunk-at-new-york-air-brake/SP-CAAAMBV

THANK YOU

splunk>