Predicting the Result of Incidents with Big Data and Data Science

Michael Fish, Pentaho Sales Engineer, Hitachi Vantara
Caio Moreno, Pentaho Consultant, Hitachi Vantara
Anya Rumyantseva, Pentaho Data Scientist, Hitachi Vantara

HITACHI
Inspire the Next

PentahoWorld

Agenda

Solving Network Rail challenges using Pentaho and Data Science techniques

- Network Rail overview

- Challenges

- Implemented approach

- Demo

- Future vision

Who Is the Prospect?

Network Rail is the owner and infrastructure manager of the rail network in England, Scotland and Wales.

- Manage the Infrastructure that all trains run on
- Their job to get as many services running on the network as possible
- If services are late or don't run and it's Network Rails fault; they are heavily fined

What's the challenge made up from?

- 20 thousand miles or track
- 40 thousand bridges and tunnels
- More than 25 hundred stations
- Bridges, Level Crossings, embankments etc.
- Freight Services
- Financial performance metric is -£200m below target*!

* According to the latest Quarterly report (Q3 2016/2017)

CHALLENGES AHEAD

Increasing demand on Britain's railway

Busiest rail network in the world

Oldest rail network in the world

Train delays are frustrating...

...and expensive
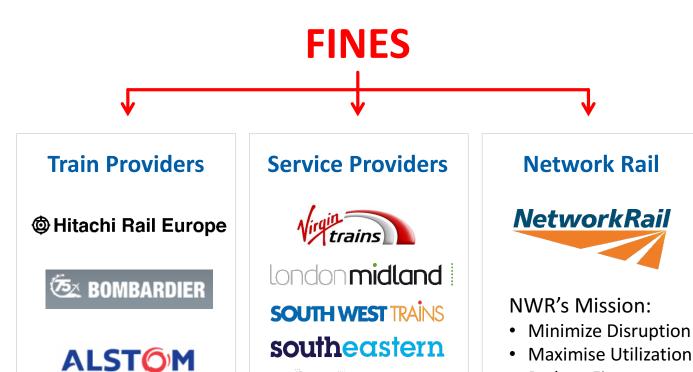
How Are Delays Handled?

- Who's fault was the delay?

- What was the cost of the delay?

- Who has to pay for the delay?
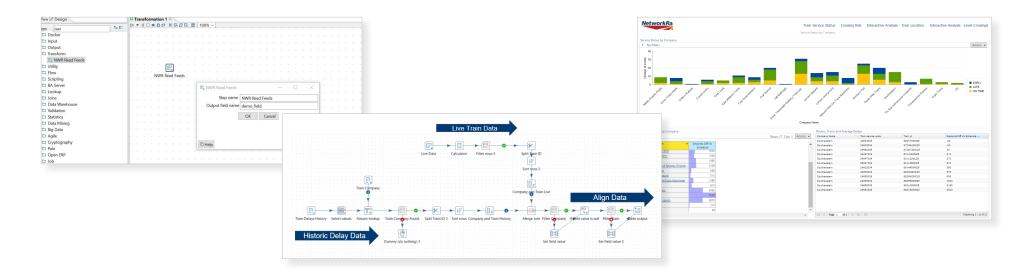
First Step: Delay Attribution

FINES

Train Providers	Service Providers	Network Rail
Hitachi Rail Europe	Virgin trains	NetworkRail
BOMBARDIER	london midland	
ALSTOM	SOUTH WEST TRAINS	
	southeastern	
	ScotRail SCOTLAND'S RAILWAY	

NWR's Mission:
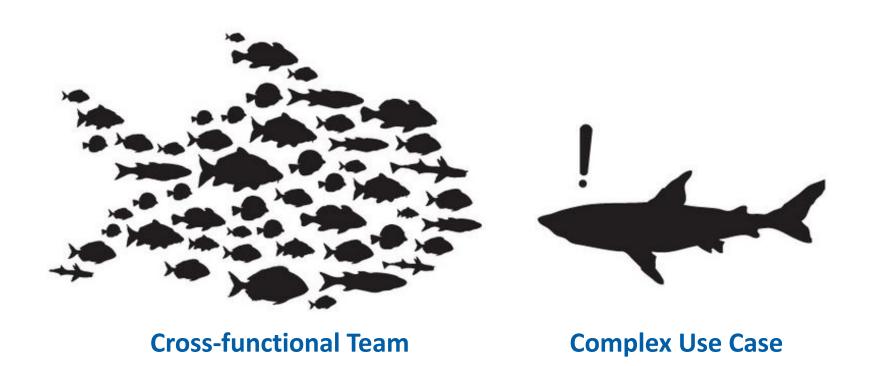- Minimize Disruption
- Maximise Utilization
- Reduce Fines
- Increase Profits

Our Approach

- Develop connectors to Network Rail data feeds in Pentaho

- Build Pentaho dashboards for delays monitoring

- Apply machine learning algorithms to the train delays data

Our Approach

Cross-functional Team

Complex Use Case

PentahoWorld

How It Worked

Pentaho Platform
Engineer solution to collect Data, blend, prepare and visualize results
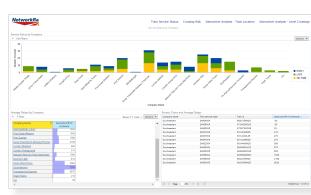
Services
To connect to NWR data Service – Custom Plugin

Data Science
Analyse historic delays to model delay attribution

Results!

Demo Agenda

- Connecting to Network Rail Live Data Feeds

- Analytics on Live Train Movement Data

- Pentaho Dashboards for Rail Network Condition Monitoring

- Predicting Incident Attribution: Machine Learning Orchestration in Pentaho

Summary

What we covered today:

- The Challenges faced by the rail network

- This isn't just a BI problem, it's a data engineering and data science problem

- Successfully realized high value in predicting the attribution of delays