Tackling Fleet Test Data with MATLAB
An Automotive Fleet Data Case Study

Will Wilson

Application Engineer

Agenda

- Motivation & Common Understanding

- Case Study: Automotive Fleet Data Analytics

- Key Takeaways

Why Analyze Vehicle Fleet Data?

- Understand real-world behavior (spec vs. actual usage)

- Make decisions with **<u>all</u>** of your data

- Enable Domain Experts to validate component or system behavior ***in situ***

- Identify and open up new market / product opportunities

BMW Uses Machine Learning to Detect Oversteering

Challenge

Develop automated software for detecting oversteering, an unsafe condition in which rear tires lose their grip during a turn

Solution

Use MATLAB to develop, train, and evaluate a variety of supervised machine learning classifier types, including KNN, SVM, and decision trees

Results

- Oversteering identified with greater than 98% accuracy
- Multiple machine learning classifiers trained automatically
- Code generated and deployed to an ECU for real-time, in-vehicle testing

A BMW M4 oversteering on a test track.

"Working in MATLAB, we developed a supervised machine learning model as a proof of concept. Despite having little previous experience with machine learning, in just three weeks we completed a working ECU prototype capable of detecting oversteering with over 98% accuracy."

- Tobias Freudling, BMW Group

Link to article

Suzuki Motor Corporation leverages Machine Learning as part of a MATLAB based workflow to optimize control parameters

Optimize control parameters to achieve "Drivability feel"

- Need to standardize

Challenges

- Drivability was quantified based on individual experience and sensory feel
- Model creation required lots of trial and error, need to iterate quickly was important
- Computing time
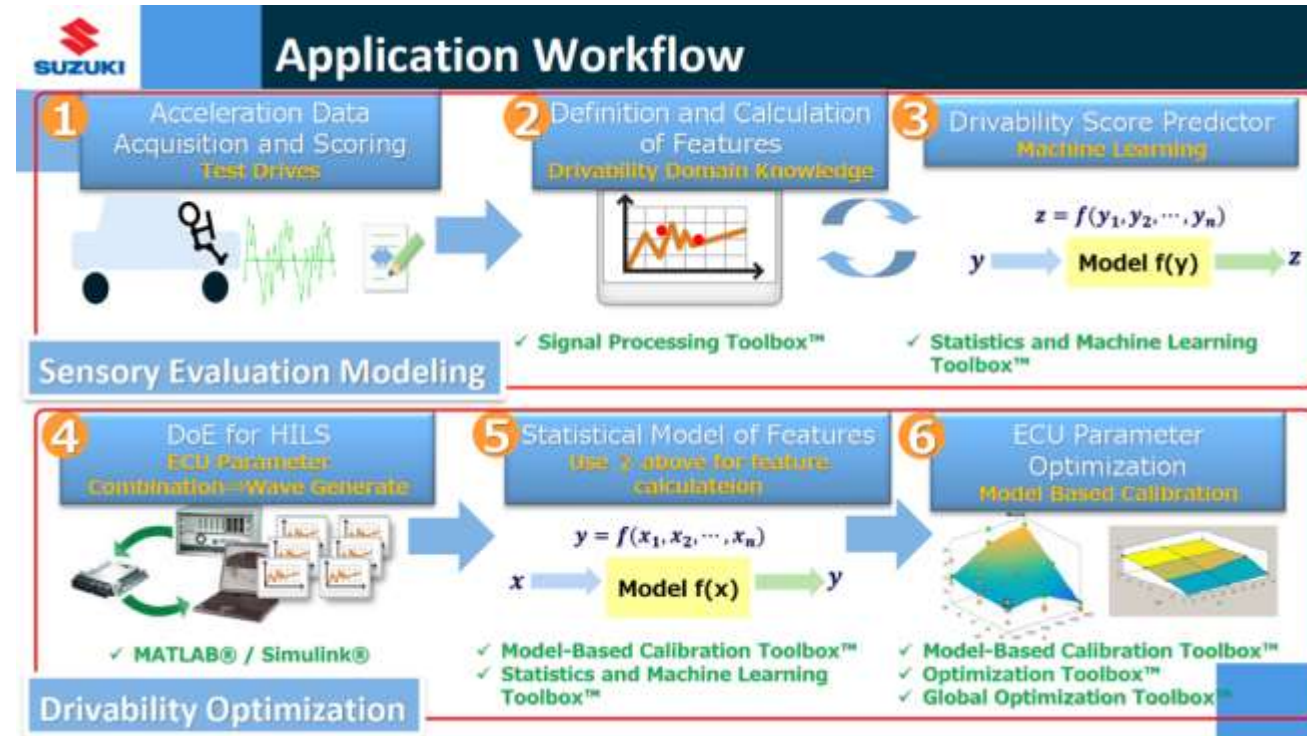
Source: „Machine Learning Based Sensory Evaluation Modeling for Vehicle Driveability Optimization"
Yoshinao Okajima, Suzuki Motor Corporation
MATLAB EXPO Japan, 2017

Workflows Enable Domain Experts to Develop Fleet Analytics
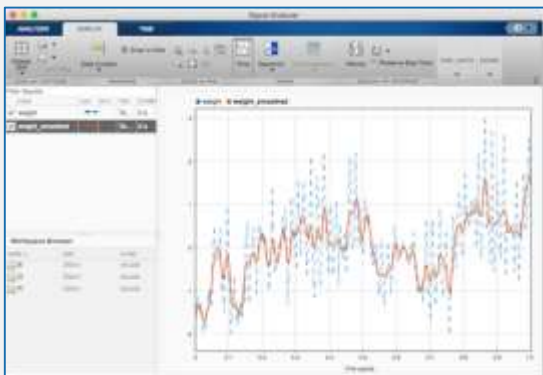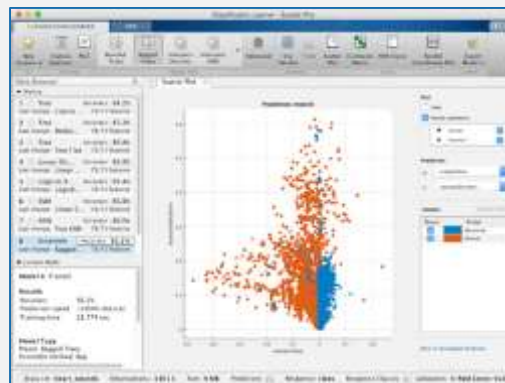## Complete Workflows

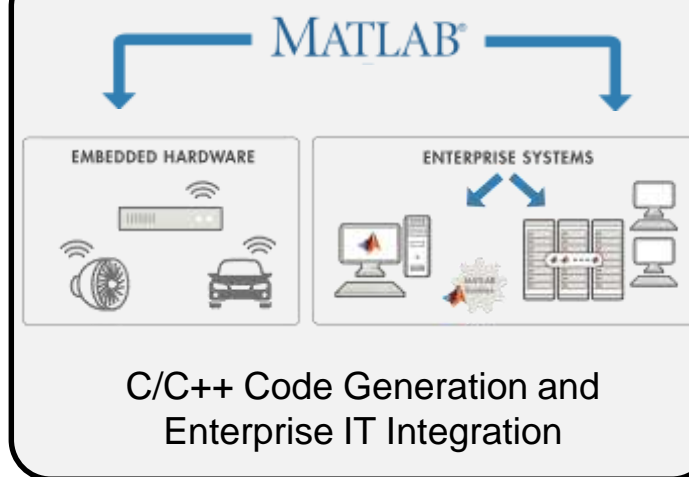**Data Access** → **Preprocessing** → **Feature Engineering** → **Model Training** → **Model Tuning** → **Deployment**
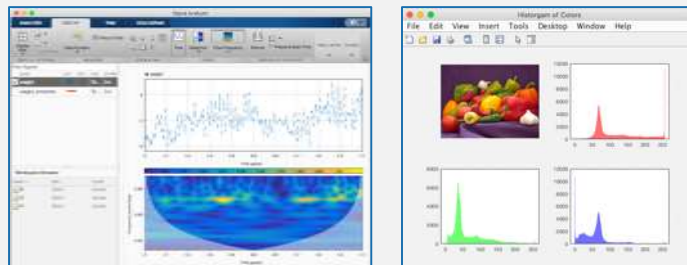
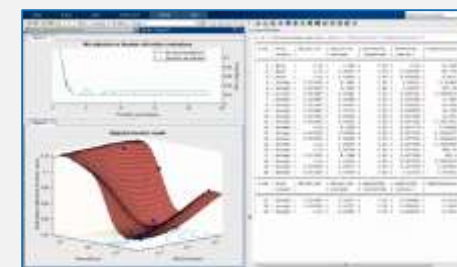Datatypes and tools for missing data, outliers, time-alignment, etc.

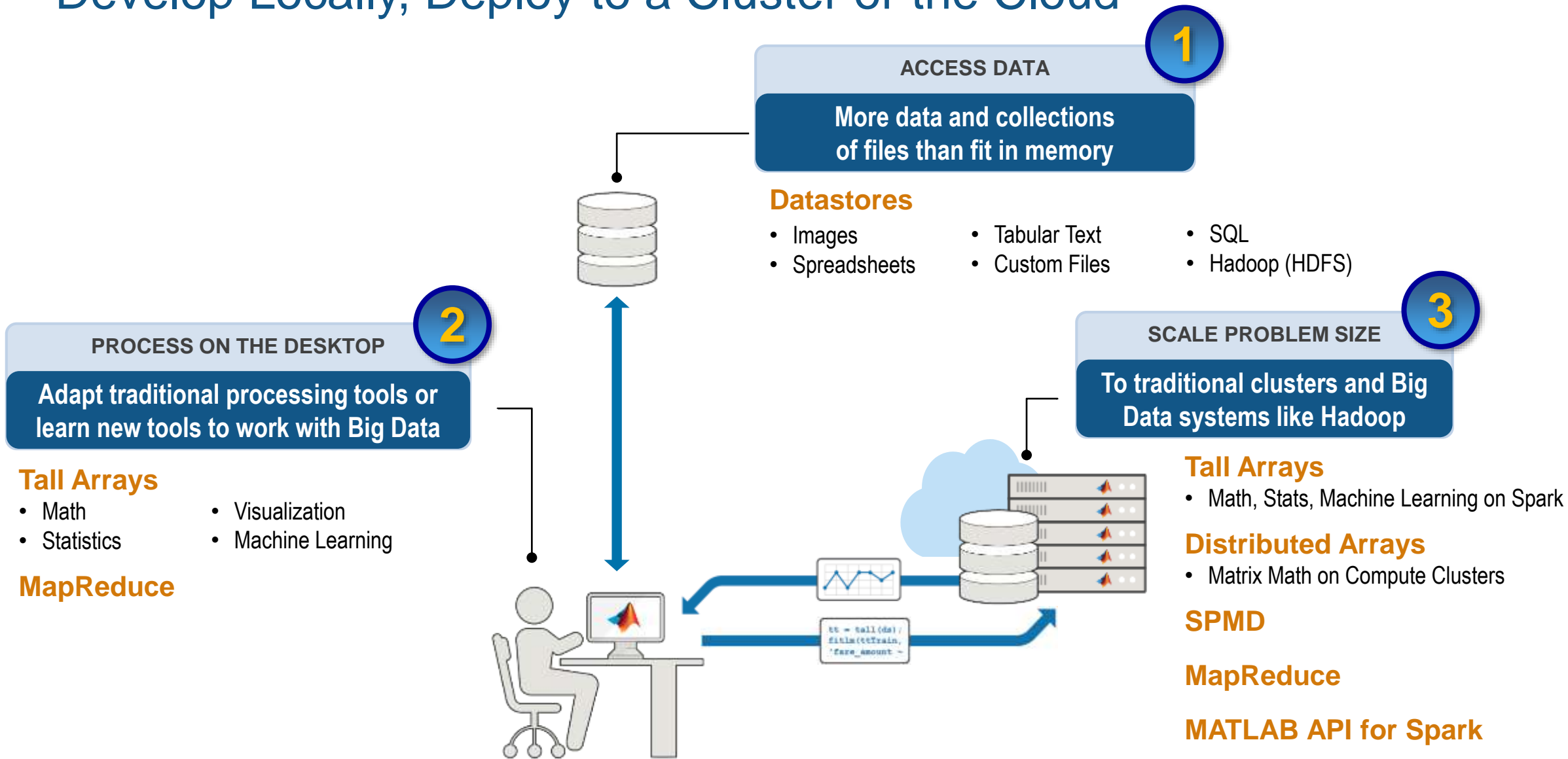Text files, spreadsheets, databases, binary files, data feeds, web, cloud storage

Machine Learning apps

Domain-specific techniques for Signals, Images, Video, Audio, and Text

C/C++ Code Generation and Enterprise IT Integration

Automated Parameter Tuning

Develop Locally, Deploy to a Cluster or the Cloud

1

ACCESS DATA

More data and collections of files than fit in memory

Datastores

- Images
- Spreadsheets
- Tabular Text
- Custom Files
- SQL
- Hadoop (HDFS)

2

PROCESS ON THE DESKTOP

Adapt traditional processing tools or learn new tools to work with Big Data

Tall Arrays

- Math
- Statistics
- Visualization
- Machine Learning

MapReduce

3

SCALE PROBLEM SIZE

To traditional clusters and Big Data systems like Hadoop

Tall Arrays

- Math, Stats, Machine Learning on Spark

Distributed Arrays

- Matrix Math on Compute Clusters

SPMD

MapReduce

MATLAB API for Spark

Challenges with Analyzing Automotive Fleet Data

| Challenge	How does MATLAB Help?
Time	Out of the box, MATLAB R2018a is 2x faster than MATLAB R2015a (or older)
Too much data / very large individual files	`datastores, tall Arrays, MATLAB MapReduce, Parallel Computing`
Time Series Data (*with different sample rates*)	`timetable, retime, synchronize`
Messy / Missing / Incomplete Data	Extensive pre-processing functionality
File Formats	Native support for parquet and MDF file formats, custom datastores
Integrate data-oriented algorithms with classical controls	MATLAB & Simulink in a single platform

Lesson's Learned About Automotive Vehicle Test Fleets
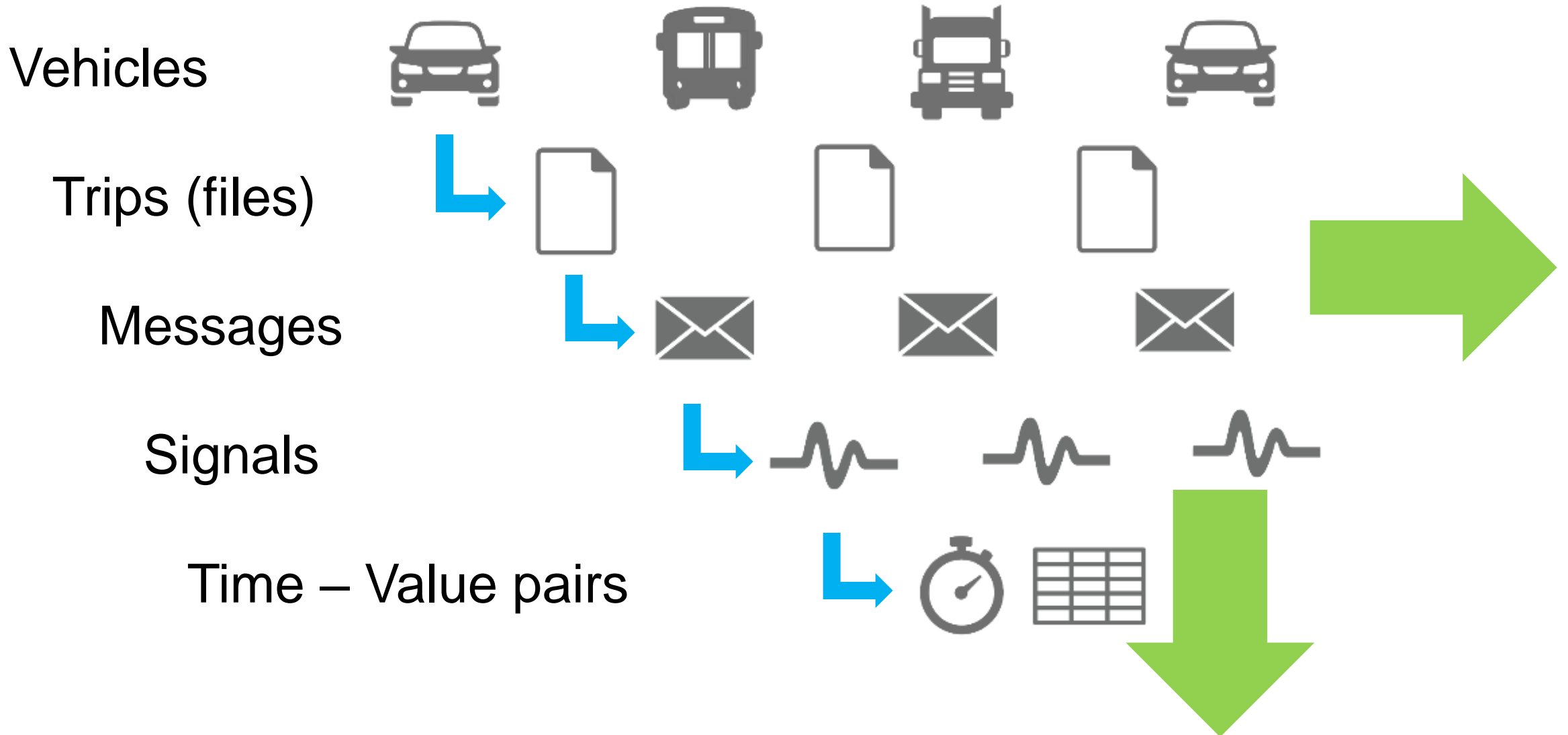
Vehicles

Trips (files)

Messages

Signals

Time – Value pairs

Agenda

- Motivation & Common Understanding

- **Case Study: Automotive Fleet Data Analytics**

- Key Takeaways

Case Study: MathWorks Vehicle Fleet

Challenge

- Given a large set of vehicle fleet data:
  - Efficiently explore all of the available data to deepen our understanding
  - Develop and implement "event detection" functionality
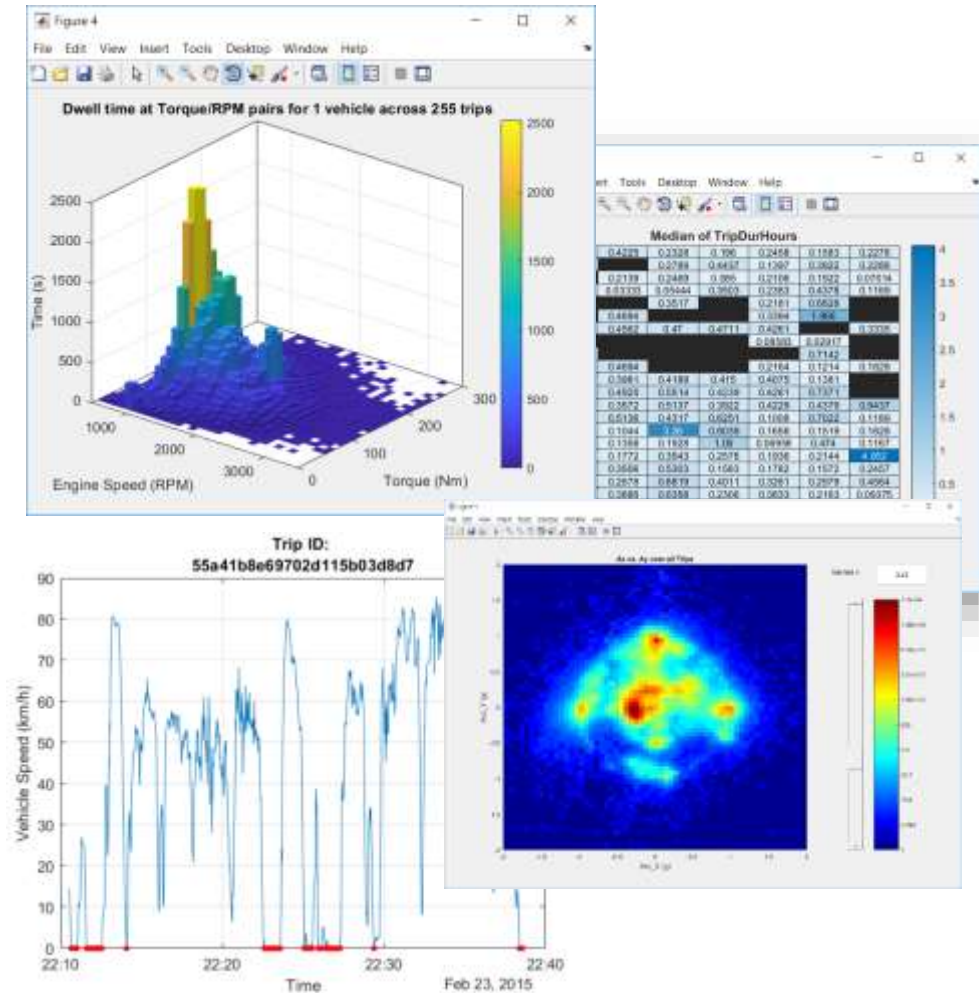 - Develop a Machine Learning model to classify driving behavior

Results

- Utilizing the MATLAB platform and `tall` arrays, each of the challenge requirements above were delivered on in a flexible, scalable way.

Demo
Parts 1 & 2

Case Study Results – Exploration & Discovery

- Worked with all of our data with `tall` arrays
  - Trip based analytics (**"for Each"**)
  - Fleet based analytics (**"for All"**)

- Extracted performance metrics & created summary visualizations

- Applied an "Event Detection" function to all of our data to find specific information

Case Study: MathWorks Vehicle Fleet

Challenge

- Given a large set of vehicle fleet data:
    - Efficiently explore all of the available data to deepen our understanding
    - Develop and implement "event detection" functionality
   - Develop a Machine Learning model to classify driving behavior

Results

- Utilizing the MATLAB platform and `tall` arrays, each of the challenge requirements above were delivered on in a flexible, scalable way.

Demo
Part 3

Case Study Results – Model & Predict

- Interactively explored classification algorithms with the Classification Learner App

- Investigated ways to incrementally improve model performance

- Generated c-code from our MATLAB based Machine Learning model to enable faster prediction time
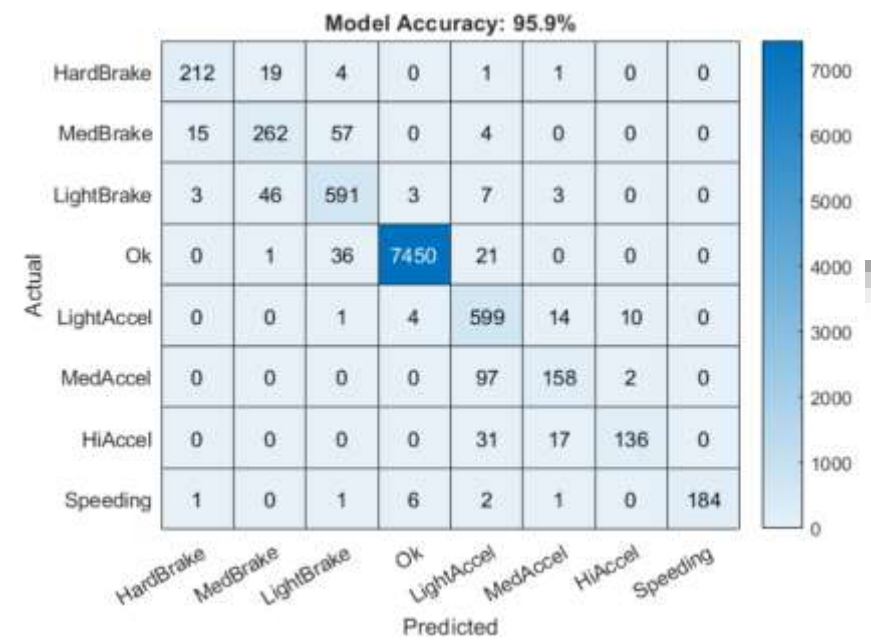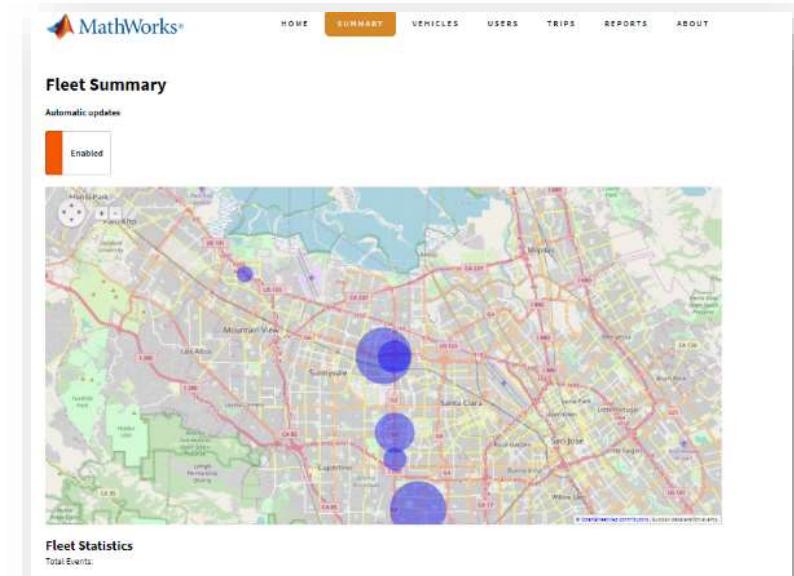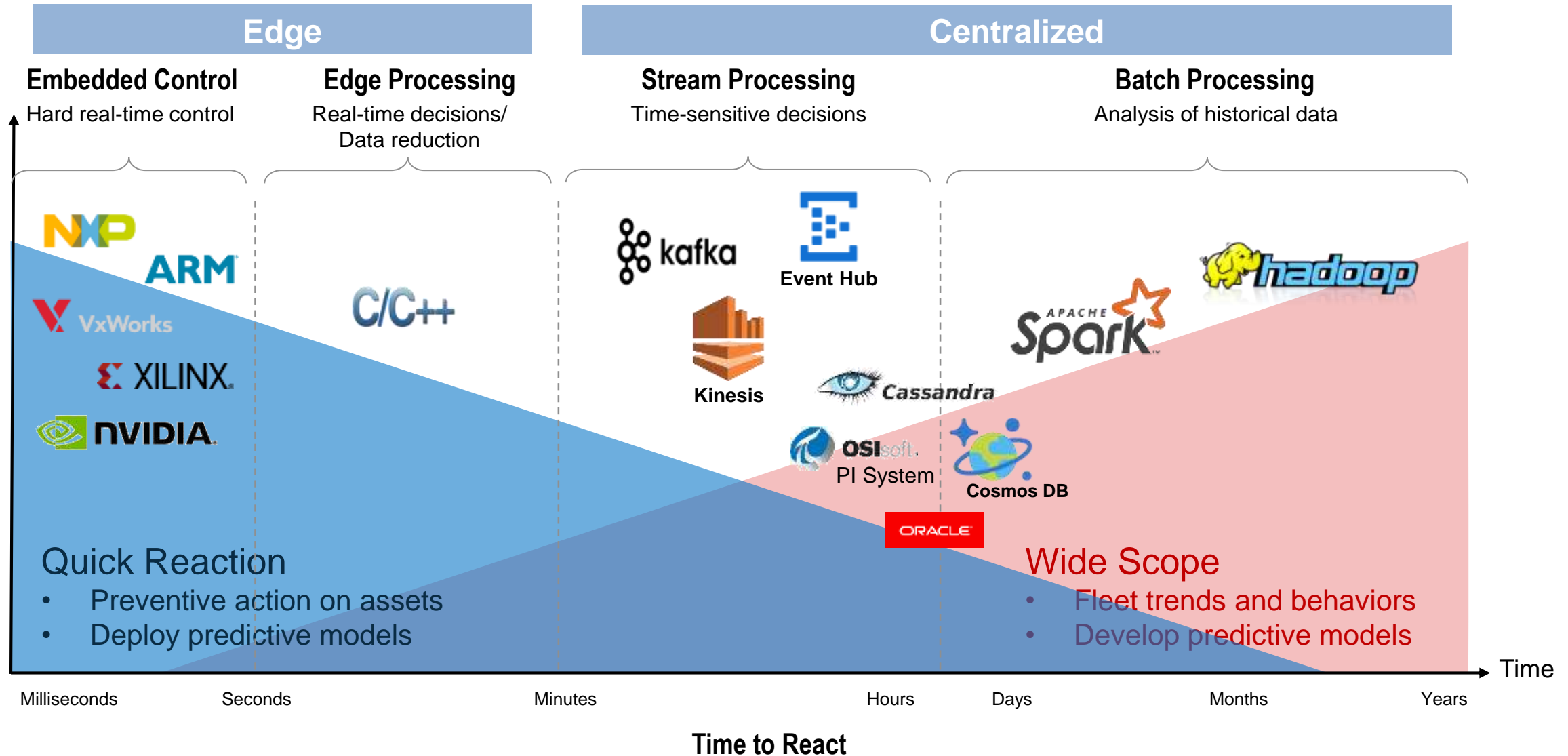
Workflow Pattern / Recipe

- Access out of memory data

- Aggregate, summarize, & visualize

- Develop functions for event detection and calculation

- Leverage table & timetable functions

- Apply functions to all of your data

- `datastore & tall`

- `table, histogram, heatmap, boxplot, binscatter`

- `Normal MATLAB code`

- `fillmissing, filloutliers, synchronize, retime, etc`

- `matlab.tall.transform & matlab.tall.reduce`

Agenda

- Motivation & Common Understanding

- Case Study: Automotive Fleet Data Analytics

- **Key Takeaways**

MATLAB Integrates Across the Spectrum of Processing Needs

Edge | **Centralized**

Embedded Control
Hard real-time control

Edge Processing
Real-time decisions/
Data reduction

Stream Processing
Time-sensitive decisions

Batch Processing
Analysis of historical data

Quick Reaction
- Preventive action on assets
- Deploy predictive models

Wide Scope
- Fleet trends and behaviors
- Develop predictive models

Time

Milliseconds   Seconds   Minutes   Hours   Days   Months   Years

Time to React

19

Key Takeaways / Call To Action

- Working with fleet test data is hard. The MATLAB platform enables **Domain Experts** to implement efficient **Data Analytics** ranging from simple exploratory work to deployment of full featured **Machine Learning** Models.

- Make sure you are thinking about the "**Everything Else**" required to achieve your goals. Data ingestion, cleanup, and deployment are just as important as the modeling / Machine Learning step in the process.

- Connect with me to discuss fleet analytics: wwilson@mathworks.com