Fleet Analytics with MATLAB

 TOOLING TO **work with** & questions that can be **asked of** fleet data

Will Wilson
Application Engineer
Have you ever wondered how...
Event detection as an enabling workflow

- Different factors affect how a particular driver drives?

- To study and understand real-world system performance?
  - Things like: Fuel economy, Emissions, ADAS features, Vehicle dynamics, Ride and handling, Prognostics, or Durability?

- To iterate through your fleet data (Terabytes perhaps) to extract relevant time slices for further study or analysis?
Event Detection in “Big Data” - the Big Idea
From Macro to Micro

- Huge sets of time history data where some sections of it are interesting. How do you get the important “time-slices” and ignore the rest?

~ 8 months of driving
(could be hundreds of trips)

3.5 minutes of something important to analyze
Event Detection in the context of Fleet Data
4 requirements for success

1. That you can explicitly define what an “Event” is
   – Could be anything from a simple logical statement to a threshold to a cutting-edge AI algorithm

2. If you can detect it, that you can manage that knowledge
   – This is all about keeping track in a way that enables analytics

3. That you can index into timeseries signals based on event time(s)

4. That you have sufficient compute power and scale to do the work
MATLAB Code Deployed as Spark on Hadoop

Short demo
What did we just see?
Multiple jobs chained together

- **Event Detection**
  - Find the things we care about (in time)
  - Out of this step is a MATLAB table with the critical information

- **Signal Extraction based on Event Detection**
  - Use what we learned from event detection to select (i.e., extract) any other signal data in that set of time ranges.
  - Output of this step is also a MATLAB table that contains a collection of tables
Visually…

Use one or more signals to “find” the time slices you care about, record where they occur

Step 1
Event Detection
Business value of Fleet Analytics

*Enabling understanding, saving time, saving money*

- Faster time to insights (Design)
  - Inform future designs (over-design vs. spec)
  - Uncover unanticipated failure or operating modes
  - Quantify real-world usage measurements

- New business opportunities in the Mobility Age
  - Uber, Automation, Services

- Warranty / brand reputation
  - Earlier in-field detection of warranty issues
  - Design better test plans
What is a Fleet?

- A fleet is a collection (group) of capital resources that generate operational or measured data you want or need to act on.

**Automotive**
- Vehicles
- Engines
- Controllers

**Manufacturing**
- Pick & Place machines
- Welding robots
- Material handling systems

**Energy**
- Wind Turbines
- Solar Panels
- Generators

**Agriculture**
- Harvesters
- Tractors
- Mining

**Healthcare**
- Surgical tools
- Wearables
- Digital health equipment

**Infrastructure**
- Charging stations
- Parking spaces
- Electronic toll collection
What is Fleet Data?
Fleet Data is not transactional data

<table>
<thead>
<tr>
<th>Business Data (i.e., Transactional Data)</th>
<th>Engineering Data (i.e., Time-series data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical example</td>
<td>Sensor data logged vs. time</td>
</tr>
<tr>
<td>Web logs, query results, etc</td>
<td>Generally not text: .dat, .mdf, binary, .mat, .blf</td>
</tr>
<tr>
<td>Common formats</td>
<td></td>
</tr>
<tr>
<td>Text, .csv, Excel, JSON</td>
<td></td>
</tr>
<tr>
<td>Storage Location</td>
<td>Files</td>
</tr>
<tr>
<td>Database</td>
<td></td>
</tr>
<tr>
<td>Partitioning</td>
<td>Not easily partitionable</td>
</tr>
<tr>
<td>Easily partitionable</td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>Engineering, Manufacturing, etc</td>
</tr>
<tr>
<td>Finance, Marketing, etc</td>
<td></td>
</tr>
<tr>
<td>Industries</td>
<td>Auto, Agriculture, Aerospace, Defense, Manufacturing, Finance, Medical</td>
</tr>
<tr>
<td>All</td>
<td></td>
</tr>
</tbody>
</table>
What are Fleet Analytics?

- Fleet analytics are the **results** of questions you ask of your data **that enable business decisions**. These questions typically fall into 2 broad categories:
  
  **For each** *(trip, day, serial #, customer, etc)* in the fleet data set, calculate some Key Performance Indicator *(KPI*) given parameters XYZ".

  **Across All** *(data) in the fleet data set, calculate descriptive statistics of specific variables (min, max, median, count, etc.) to summarize and visualize (histograms).*
**Big Data file-based workflows**

*Note: tall is parallel by default but can be configured to run in serial if necessary.*

```matlab
function events = detectEvents(t)
    %<your own custom MATLAB code here>
end

datastore

```

### Question Type

- **“For Each”**
- **“Across All”**

```matlab
tds = transform(ds, @detectEvents);
TT = tall(tds);
EventsSummary = gather(TT);

MedianValSigABC = gather(median(TT.SignalABC));
```
Key Aspects of Fleet Data

Fleet Analytics requires managing data at 2 levels

**Metadata**

*(Describing information)*

- This is the information that you need in order to report / summarize.
- Examples include things like:
  - VIN, serial #, unit #, etc
  - Calibration / Software ID
  - Test # / Test Case

**Timeseries data**

*(Measurement information)*

- This is the actual engineering data we need to do calculations. It is typically collections of sensor readings with respect to time.
- Examples include:
  - Vehicle speed
  - Engine RPM
  - Etc…
Fleet Analytics – a Retrospective (1/4)
Lesson’s Learned from several years of working with customers on this topic

- Your analytics will evolve, prepare for this now
  - MATLAB Projects, Source Control, good design
Lesson’s Learned from several years of working with customers on this topic

- Strive for efficient (and flexible) data preprocessing
  - Use built in functions, datastore writeall, file format(s)

- Timeseries work (timetable)
  - Resample or aggregate data in timetable, and resolve duplicate or irregular times
  - Synchronize timetables to common time vector, and resample or aggregate data from input timetables

- 30+ functions to help you resolve problems like:
  - Missing Data and Outliers
  - Detecting Change Points and Local Extrema
  - Smoothing and Detrending Data
  - Normalizing and Scaling Data
  - Grouping and Binning Data

Native support for reading:
- MDF / MF4
- blf
- parquet
- text
- .mat
Fleet Analytics – a Retrospective (3/4)
Lesson’s Learned from several years of working with customers on this topic

- As your data grows, so must your compute

**Desktop Compute**
- Single machine
- Typically tens of cores
- Leverage GPU or Multi-Core CPU

**High Performance Computing (HPC)**
- Multiple machines
- Typically tens to hundreds of cores
- Leverage GPU or Multi-Core CPU

**Reference architectures**
- **MATLAB Parallel Server**
- **MATLAB Compiler**

“Big Data” Compute
- Multiple machines
- Typically hundreds of cores

**Bring the data to the compute**
- **Parallel Computing Toolbox**
- **MATLAB Parallel Server**
- **Spark**
- **Hadoop**

**GitHub**

**Bring the compute to the data**
Lesson’s Learned from several years of working with customers on this topic

- Careful partitioning of your data is key to performant analytics
  - Begin with the end in mind. What will you ask of your data?
  - Organize your data with what you know now, adapt as you learn
  - Separate / group your files by what is important to you (VIN, SW ID, calibration, etc.)
  - Implement descriptive (and consistent) folder and file names as much as possible
The MathWorks Platform
Comprehensive end-to-end solution for Fleet Analytics and AI

Access and Explore Data
- Operational Technology
  - OSISoft PI

Preprocess Data
- Data Sources
  - Streaming data
    - Apache Kafka
    - Azure EventHub
  - Data stores
    - AWS S3
    - Azure BLOB
  - Files
    - Avro
    - Parquet

Analyze Data
- Big Data
  - hadoop
  - Apache Spark
  - cloudera
  - databricks

Develop Algorithms & Models
- MATLAB Parallel Server
  - Training, simulation, optimization

Operationalize Models
- CI / CD
  - Azure DevOps
  - Jenkins

MATLAB Platform
- MATLAB
  - Data exploration, preprocessing, algorithm development

Domain specific toolboxes

Web App
- MATLAB Web App Server
  - RESTful API

Tableau
- Spotfire
- PowerBI
- Qlik

Enterprise App
- MATLAB Excel
- Java
- C/C++
- .NET,dll
- Python
- cloudera
- databricks
- Apache Kafka
- Azure EventHub
- OSISoft PI
- AWS S3
- Azure BLOB
- Avro
- Parquet

Sharing, deployment, integration
Key Takeaways / Call To Action

- Think about the questions you need to answer
  - “For Each” vs. “Across All” (maybe even both)

- Use the right datatypes and tools to keep your code flexible
  - tables, timetables, MATLAB Projects, GIT, etc

- Determine what role parallel computing will play in your workflow
  - It’s not if, it is when you will need to scale

- We are here to help! Engage with your Account Team to learn more about our 2-hour Fleet Analytics Seminar for your team.
MATLAB Training Courses for Data Science

Training Courses

- MATLAB Fundamentals (3 days)
- MATLAB for Data Processing and Visualization (1 day)
- Processing Big Data with MATLAB (1 day)
- Statistical Methods in MATLAB (2 days)
- Machine Learning with MATLAB (2 days)
- Signal Preprocessing and Feature Extraction with MATLAB (1 day)
- Deep Learning with MATLAB (2 days)
- Accelerating and Parallelizing MATLAB Code (2 days)

Get Started for Free

MATLAB Onramp
Get started quickly with the basics of MATLAB®.
» Details and launch

Machine Learning Onramp
An interactive introduction to practical machine learning methods for classification problems.
» Details and launch

Deep Learning Onramp
Get started with deep learning techniques to perform image recognition.
» Details and launch

Practical Data Science with MATLAB Specialization

- Exploratory Data Analysis
- Data Processing and Feature Engineering
- Predictive Modeling and Machine Learning
- Data Science Project