MATLAB EXPO 2017
Integrating MATLAB Analytics into Enterprise Applications
Agenda

- Example Problem
- Access and Preprocess Data
- Develop a Predictive Model
- Integrate Analytics with Production Systems
- Build a Dashboard to Visualize Results
Example Problem – How’s my driving?

- A group of MathWorks employees installed an OBD dongle in their car that monitors the on-board systems.
- Data is streamed to the cloud where it is aggregated and stored.
- I would like to use this data to score the driving habits of participants.
Fleet Analytics Architecture
Data Analytics Workflow

Access and Explore Data

Preprocess Data

Develop Predictive Models

Integrate Analytics with Production Systems

Files

Databases

Sensors

Working with Messy Data

Data Reduction/Transformation

Feature Extraction

Model Creation e.g. Machine Learning

Parameter Optimization

Model Validation

Desktop Apps

Enterprise Scale Systems

Embedded Devices and Hardware
Access and Preprocess Data

Production System

Analytics Development

Storage Layer

kafka

MATLAB

Algorithm Developers

Business Decisions

End Users

MATLAB EXPO 2017
## The Data: Timestamped messages with JSON encoding

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
</table>
| {  
  "vehicles_id":
   {"$oid":"55a3fd0069702d5b41000000"}
} | {  
  "time":{"$date":"2015-07-13T18:01:35.000Z"},
  "kc":1975.0, 
  "kff1225":100.65293, 
  "kff125a":110.36619,
  
} |
| {  
  "vehicles_id":
   {"$oid":"55a3fe3569702d5c5c00020"}
} | {  
  "time":{"$date":"2015-07-13T18:01:53.000Z"},
  "kc":2000.0, 
  "kff1225":109.65293, 
  "kff125a":115.36619,
  
} |
| {  
  "vehicles_id":
   "$oid":"55a4193569702d115b000001"}
} | {  
  "time":{"$date":"2015-07-12T19:04:04.000Z"},
  "kc":2200.0, 
  "kff1225":112.65293, 
  "kff125a":112.36619,
  
} |
Data Access and Preprocessing – Challenges

Challenges

- Data aggregation
  - Different sources (files, web, etc.)
  - Different types (images, text, audio, etc.)

- Data clean up
  - Poorly formatted files
  - Irregularly sampled data
  - Redundant data, outliers, missing data etc.

- Data specific processing
  - Signals: Smoothing, resampling, denoising, Wavelet transforms, etc.
  - Images: Image registration, morphological filtering, deblurring, etc.

- Dealing with out of memory data (big data)

---

Data preparation accounts for about **80%** of the work of data scientists - Forbes
Access a Sample of Data and Develop a Preprocessing Function

Preprocess data

```
t = sortrows(t);
t = rmmissing(t,'MinNumMissing',width(t)-2);
```

```
t.smoothed = movmedian(t.SpeedGPS,5);
t.D1 = [0;diff(t.SpeedGPS)];
```

Perform windowed calculations

```
[tmin,tmax] = bounds(t.time);
tnew = tmin:seconds(10):tmax;
scoresByTime = retime(t(:, 'Event'), tnew, @histcounts);
```
Develop a Predictive Model

Edge Devices

Production System

Analytics Development

Storage Layer

MDCS

MATLAB

Algorithm Developers

Business Decisions

Power BI

Qlik

Spotfire

Tableau

End Users

Business Systems

MATLAB EXPO 2017
Develop a Predictive Model in MATLAB

1. Label events
2. Represent signals
3. Train model
4. Validate model
5. Scale up

MATLAB EXPO 2017
Integrate Analytics with Production Systems

Edge Devices

Kafka Connector

Production System

MATLAB Production Server

Analytics Development

MATLAB Compiler SDK

Business Decisions

Business Systems

End Users

API Gateway

AWS Lambda

kafka

Storage Layer
A quick Intro to Stream Processing

- **Batch Processing** applies computation to a finite sized historical data set that was acquired in the past

- **Stream Processing** applies computation to an unbounded data set that is produced continuously
Stream processing exploits the fact that recent data tends to be more valuable
Streaming data is treated as an unbounded table

### Input Table

<table>
<thead>
<tr>
<th>Event Time</th>
<th>Vehicle</th>
<th>RPM</th>
<th>Torque</th>
<th>Fuel Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>18:01:10</td>
<td>55a3fd</td>
<td>1975</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>18:10:30</td>
<td>55a3fe</td>
<td>2000</td>
<td>109</td>
<td>115</td>
</tr>
<tr>
<td>18:05:20</td>
<td>55a3fd</td>
<td>1980</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>18:10:45</td>
<td>55a3fd</td>
<td>2100</td>
<td>110</td>
<td>100</td>
</tr>
<tr>
<td>18:30:10</td>
<td>55a419</td>
<td>2000</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>18:35:20</td>
<td>55a419</td>
<td>1960</td>
<td>103</td>
<td>105</td>
</tr>
<tr>
<td>18:20:40</td>
<td>55a3fe</td>
<td>1970</td>
<td>112</td>
<td>104</td>
</tr>
<tr>
<td>18:39:30</td>
<td>55a419</td>
<td>2100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>18:30:00</td>
<td>55a3fe</td>
<td>1980</td>
<td>110</td>
<td>113</td>
</tr>
<tr>
<td>18:30:50</td>
<td>55a3fe</td>
<td>2000</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

### Output Table

<table>
<thead>
<tr>
<th>Time window</th>
<th>Vehicle</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>18:00:00</td>
<td>55a3fd</td>
<td>5</td>
</tr>
<tr>
<td>18:10:00</td>
<td>55a3fe</td>
<td>…</td>
</tr>
<tr>
<td>18:10:00</td>
<td>55a419</td>
<td>…</td>
</tr>
<tr>
<td>18:20:00</td>
<td>55a3fd</td>
<td>7</td>
</tr>
<tr>
<td>18:20:00</td>
<td>55a3fe</td>
<td>3</td>
</tr>
<tr>
<td>18:20:00</td>
<td>55a419</td>
<td>…</td>
</tr>
<tr>
<td>18:30:00</td>
<td>55a3fd</td>
<td>…</td>
</tr>
<tr>
<td>18:30:00</td>
<td>55a3fe</td>
<td>4</td>
</tr>
<tr>
<td>18:30:00</td>
<td>55a419</td>
<td>…</td>
</tr>
<tr>
<td>18:40:00</td>
<td>55a3fd</td>
<td>…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

**MATLAB EXPO 2017**
Introducing MATLAB Production Server

### Data
- Databases: DynamoDB, Cosmos DB, SQL Server, mongoDB
- Cloud Storage: Azure Blob, Amazon Kinesis
- Streaming: kafka, MQTT, Azure Event Hub

### Analytics
- MATLAB Production Server
- Request Broker

### Business System
- Visualization: Tableau, Qlik, Spotfire, Power BI
- Web: Microsoft IIS, Apache Tomcat, WebSphere
- Custom App: MATLAB, Business System Analytics

### Platform
- Public Cloud: Microsoft Azure, Amazon Web Services
- Private Cloud: Rackspace, OpenStack, VMware

MATLAB EXPO 2017
Introducing MATLAB Production Server

- **Server software**
  - Manages packaged MATLAB programs and worker pool

- **MATLAB Runtime libraries**
  - Single server can use runtimes from different releases

- **RESTful JSON interface**

- **Lightweight client libraries**
  - C/C++, .NET, Python, and Java
Introducing Apache Kafka

- Kafka is a high throughput distributed messaging system
- Originally developed at LinkedIn and open sourced in 2011
- Kafka is architected as a massively scalable publish/subscribe message queue
- Well suited for large scale streaming applications
Connecting MATLAB Production Server to Apache Kafka

- Kafka client for MATLAB Production Server feeds topics to functions deployed on the server

- Configurable batch of messages passed as a MATLAB Timetable

- Each consumer process feeds one topic to a specified function

- Drive everything from a simple config file
  - No programming outside of MATLAB!

MATLAB EXPO 2017
Develop, Test, and Deploy a Stream Processing Function
Develop a Stream Processing Function in MATLAB

DEMO
Test Your Stream Processing Function on Live Data
Complete Your Application

Edge Devices

Production System

Analytics Development

Business Decisions

MATLAB EXPO 2017
Complete Your Application


MATLAB EXPO 2017
Go Live!

Fleet Summary
Automatic updates

Fleet Statistics

MATLAB EXPO 2017
Key Takeaways

- MATLAB Connects directly to your data so you can quickly design and validate algorithms
- MATLAB’s high-level language and apps enable fast design iterations
- MATLAB Production Server enables easy integration of your MATLAB algorithms with enterprise production systems
- This enables you to spend your time understanding the data and designing algorithms
Resources to learn and get started

- **Data Analytics with MATLAB**
- **MATLAB Compiler SDK**
- **MATLAB Production Server**
- **Database Toolbox**