

The background features a large blue triangle on the left, a grey trapezoid at the top right, and a blue trapezoid on the right. A 3D wireframe mesh with a color gradient from blue to yellow is positioned in the lower right. Faint white lines resembling a signal waveform are visible in the grey area, and a blue circuit-like diagram is in the bottom right.

MATLAB EXPO 2017

Integrating MATLAB Analytics into Enterprise Applications

Agenda

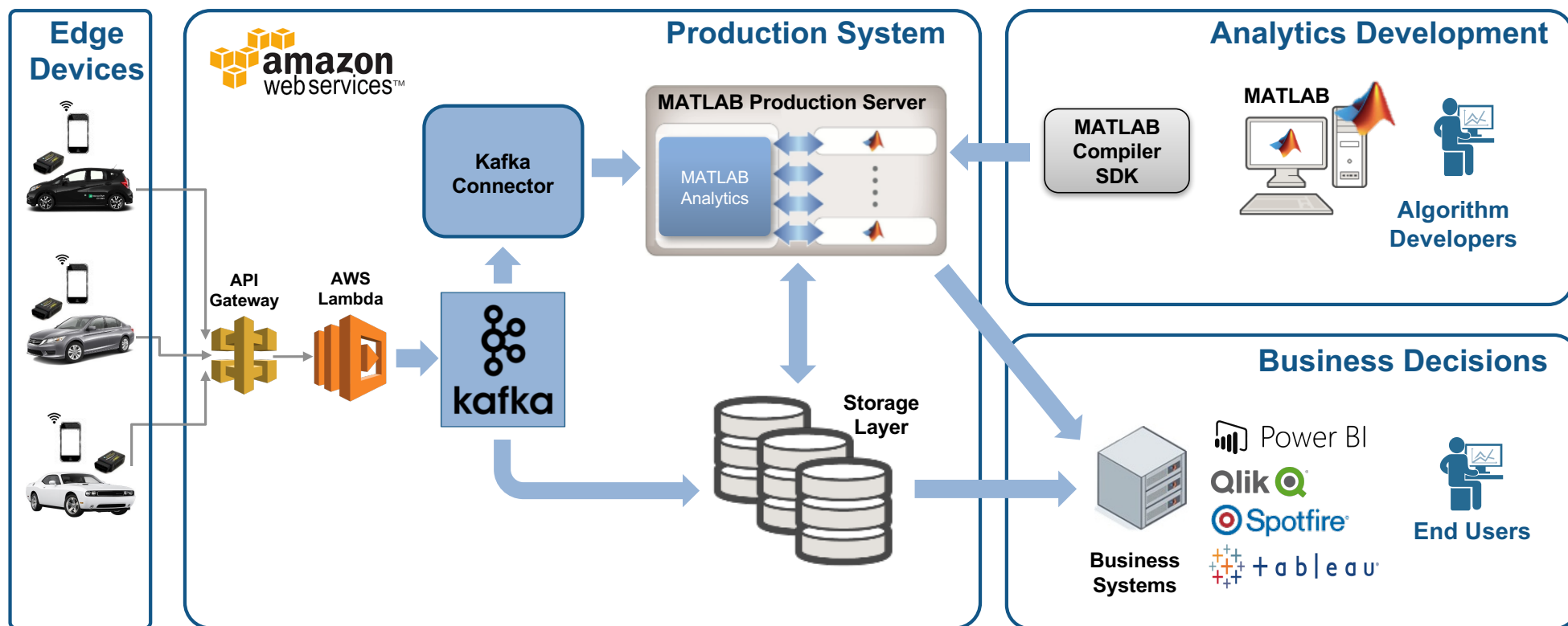
- Example Problem
- Access and Preprocess Data
- Develop a Predictive Model
- Integrate Analytics with Production Systems
- Build a Dashboard to Visualize Results

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

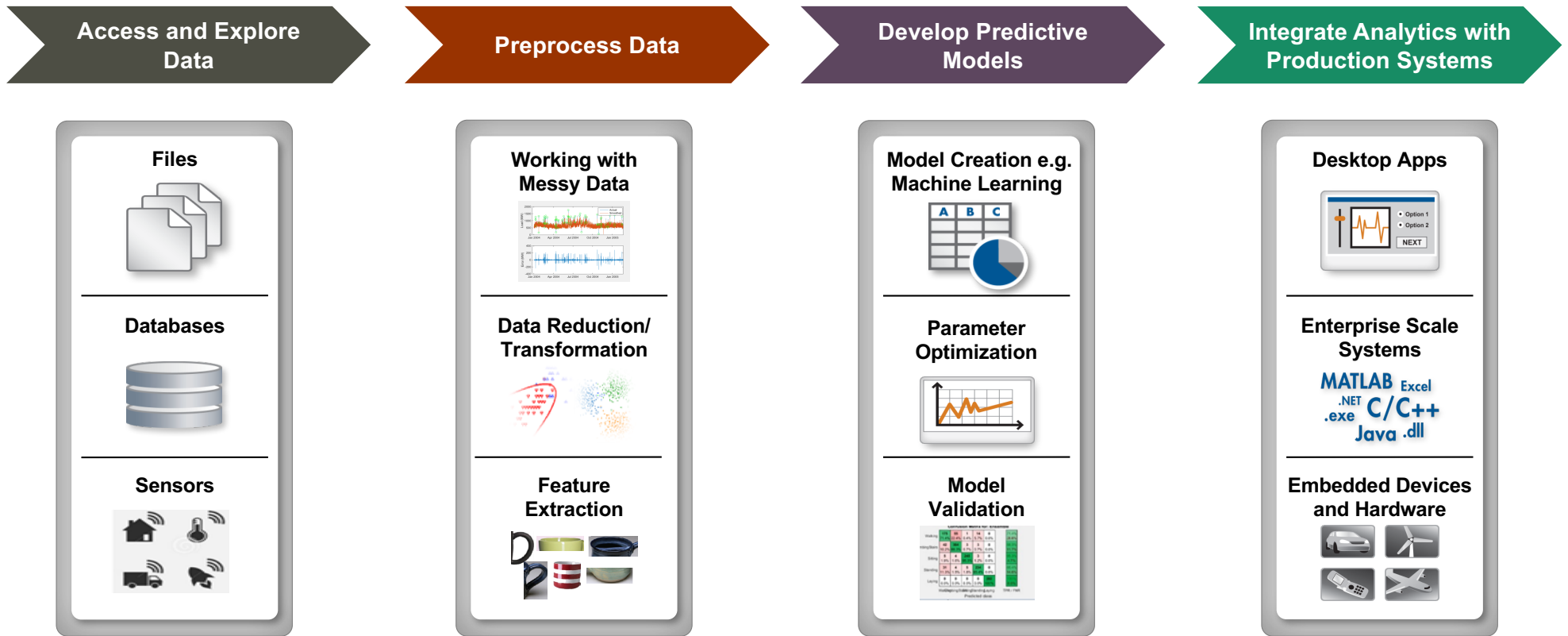


Fleet Analytics Architecture



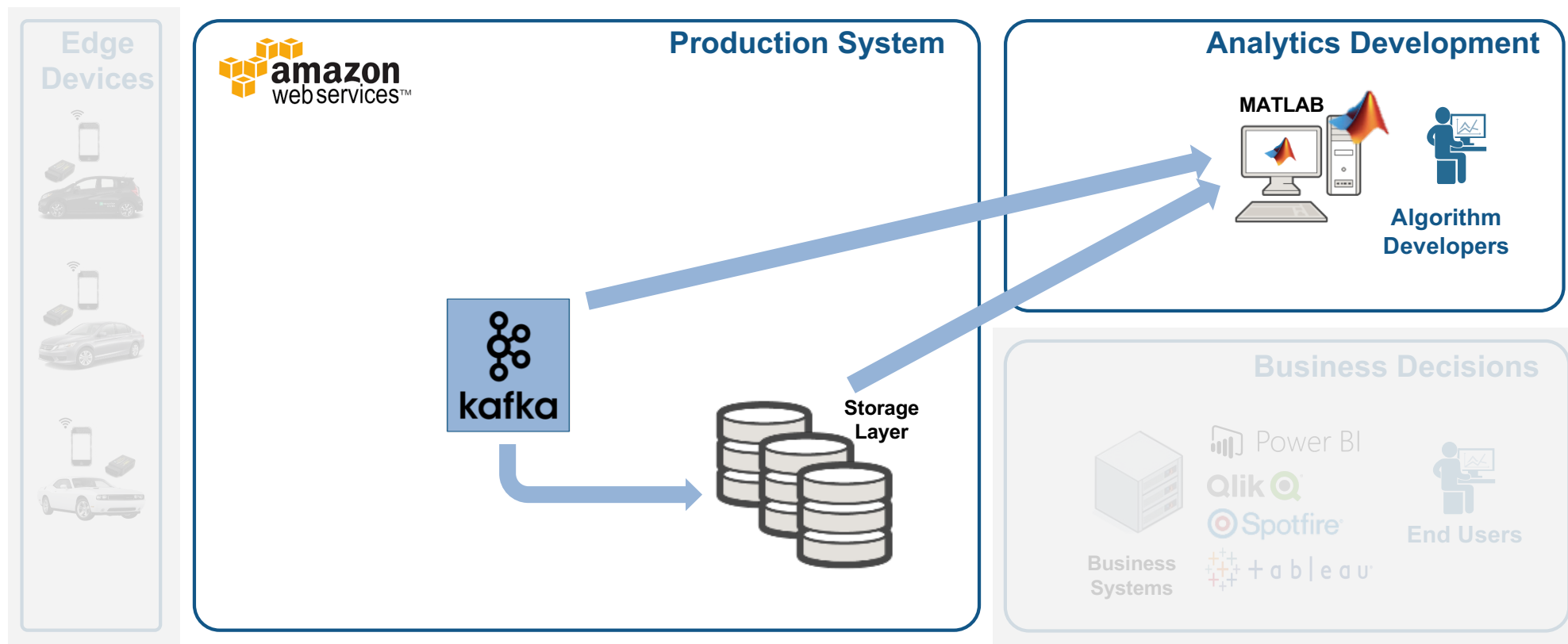
MATLAB EXPO 2017

Data Analytics Workflow






MATLAB EXPO 2017

Access and Preprocess Data



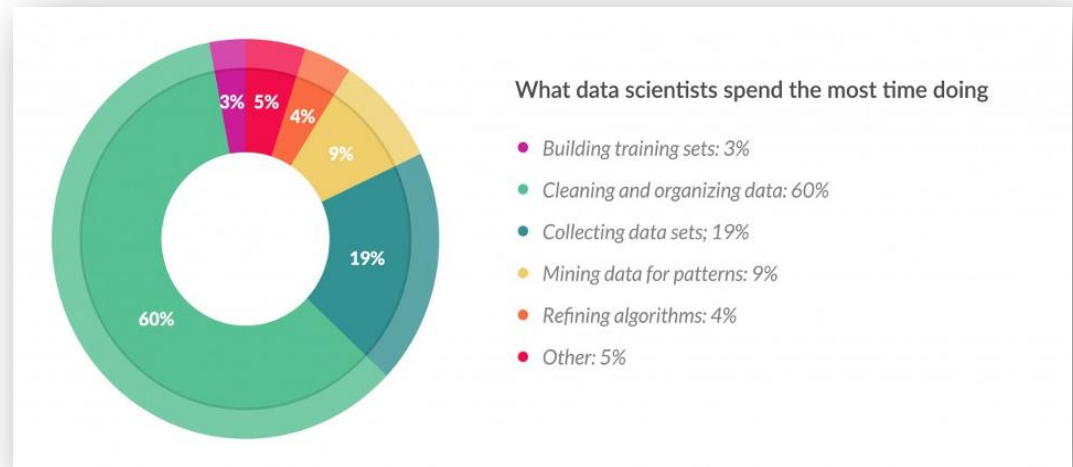
The Data: Timestamped messages with JSON encoding

	Key	Value
	<pre>{ "vehicles_id": {"\$oid":"55a3fd0069702d5b41000000"} }</pre>	<pre>{ "time" : {"\$date":"2015-07-13T18:01:35.000Z"}, "kc" : 1975.0, "kff1225" : 100.65293, "kff125a" : 110.36619, }</pre>
	<pre>{ "vehicles_id": {"\$oid":"55a3fe3569702d5c5c000020"} }</pre>	<pre>{ "time":{"\$date":"2015-07-13T18:01:53.000Z"}, "kc" : 2000.0, "kff1225" : 109.65293, "kff125a" : 115.36619, }</pre>
	<pre>{ "vehicles_id": {"\$oid":"55a4193569702d115b000001"} }</pre>	<pre>{ "time":{"\$date":"2015-07-12T19:04:04.000Z"} "kc":2200.0, "kff1225" : 112.65293, "kff125a" : 112.36619, }</pre>

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

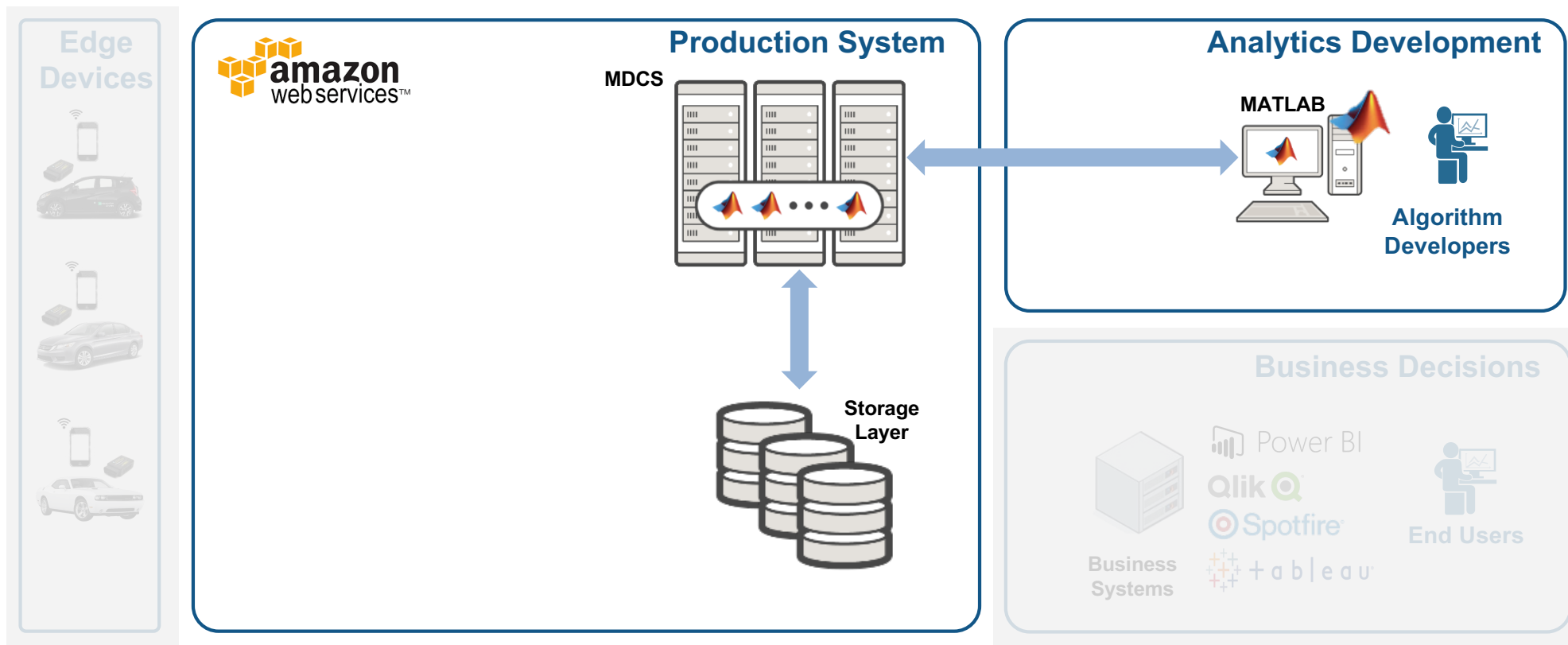
```
t = sortrows(t);  
t = rmmissing(t, 'MinNumMissing', width(t)-2);
```

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

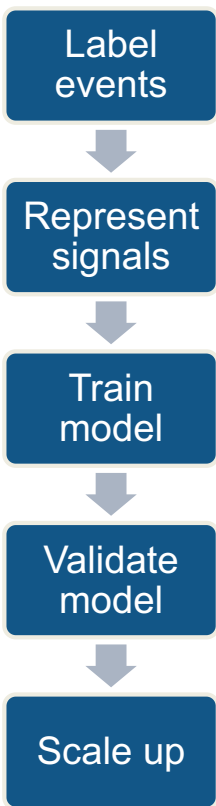
Perform windowed calculations

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

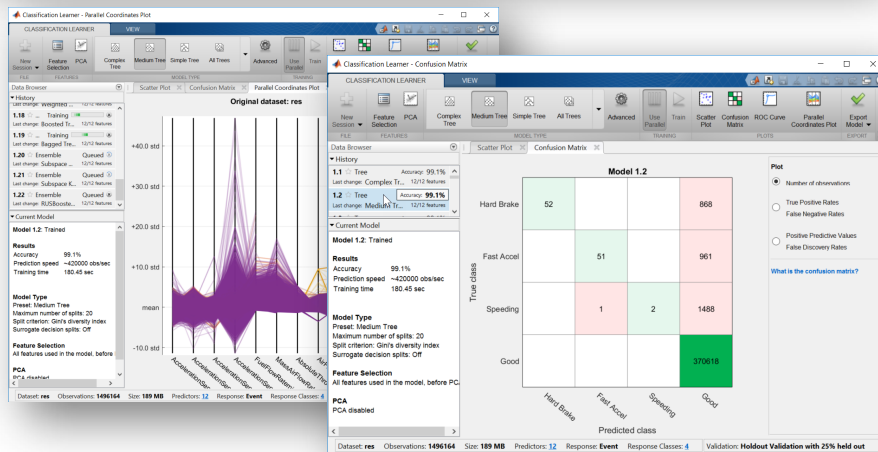
Develop a Predictive Model



Develop a Predictive Model in MATLAB



time	1 Event	2 SpeedGPS	3 AccelerationSensorXAxis	4 AccelerationSensorYAxis	5 AccelerationSensorZAxis
Mon May 11 04:03:15 UTC 2015	Hard Brake	10.8360	-0.6996	0.6014	0.205
Wed May 06 19:09:48 UTC 2015	Hard Brake	27.8280	0.1419	0.9035	-0.526
Sun May 17 17:09:19 UTC 2015	Hard Brake	6.5520	0.9986	-0.0761	-0.004
Fri Jan 16 20:38:37 UTC 2015	Hard Brake	39.6128	0.0999	0.8000	0.367
Sat May 02 14:00:37 UTC 2015	Hard Brake	61.1280	0.4006	-0.4022	0.663
Mon Apr 27 17:54:27 UTC 2015	Fast Accel	37.7640	0.1527	0.4666	0.857
Sun May 03 21:00:42 UTC 2015	Fast Accel	17.2440	1.0235	0.0815	0.304
Mon May 04 11:30:33 UTC 2015	Fast Accel	19.6560	0.1336	0.8932	-0.578
Wed May 20 10:20:55 UTC 2015	Hard Brake	22.4000	0.2050	0.0054	0.000

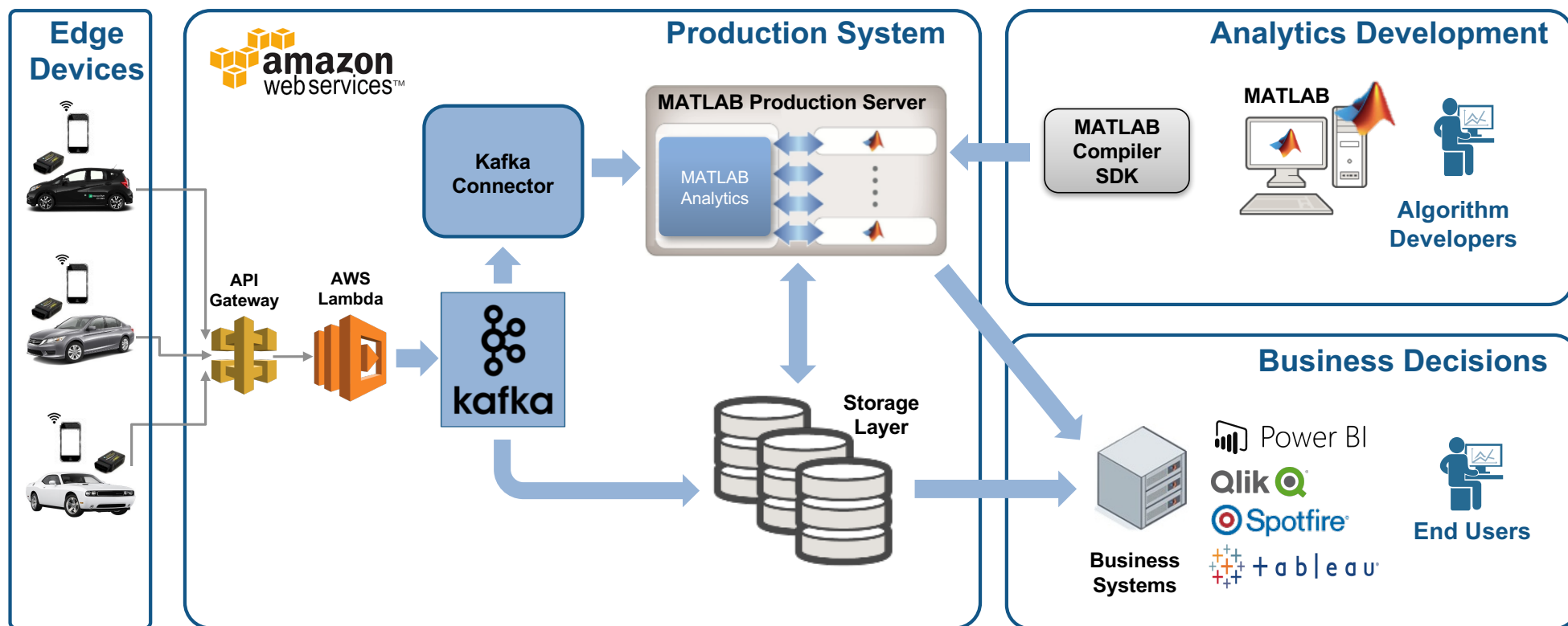


Evaluating tall expression using the Spark Cluster:

- Pass 1 of 2: Completed in 11 sec
- Pass 2 of 2: Completed in 2.333 min

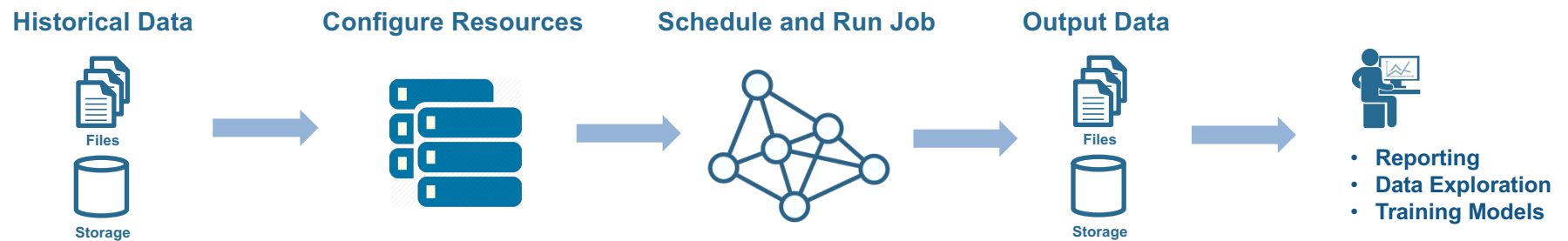
Evaluation completed in 2.6167 min

Integrate Analytics with Production Systems

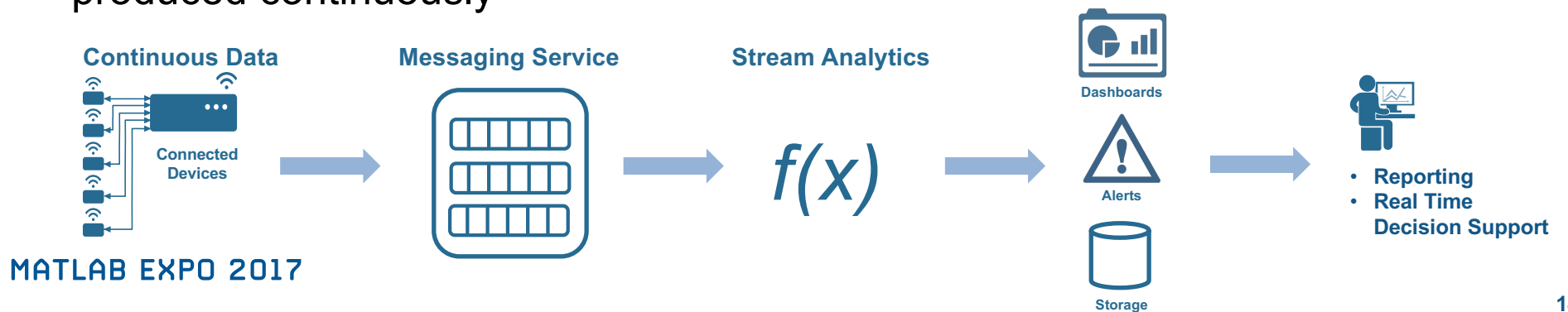


A quick Intro to Stream Processing

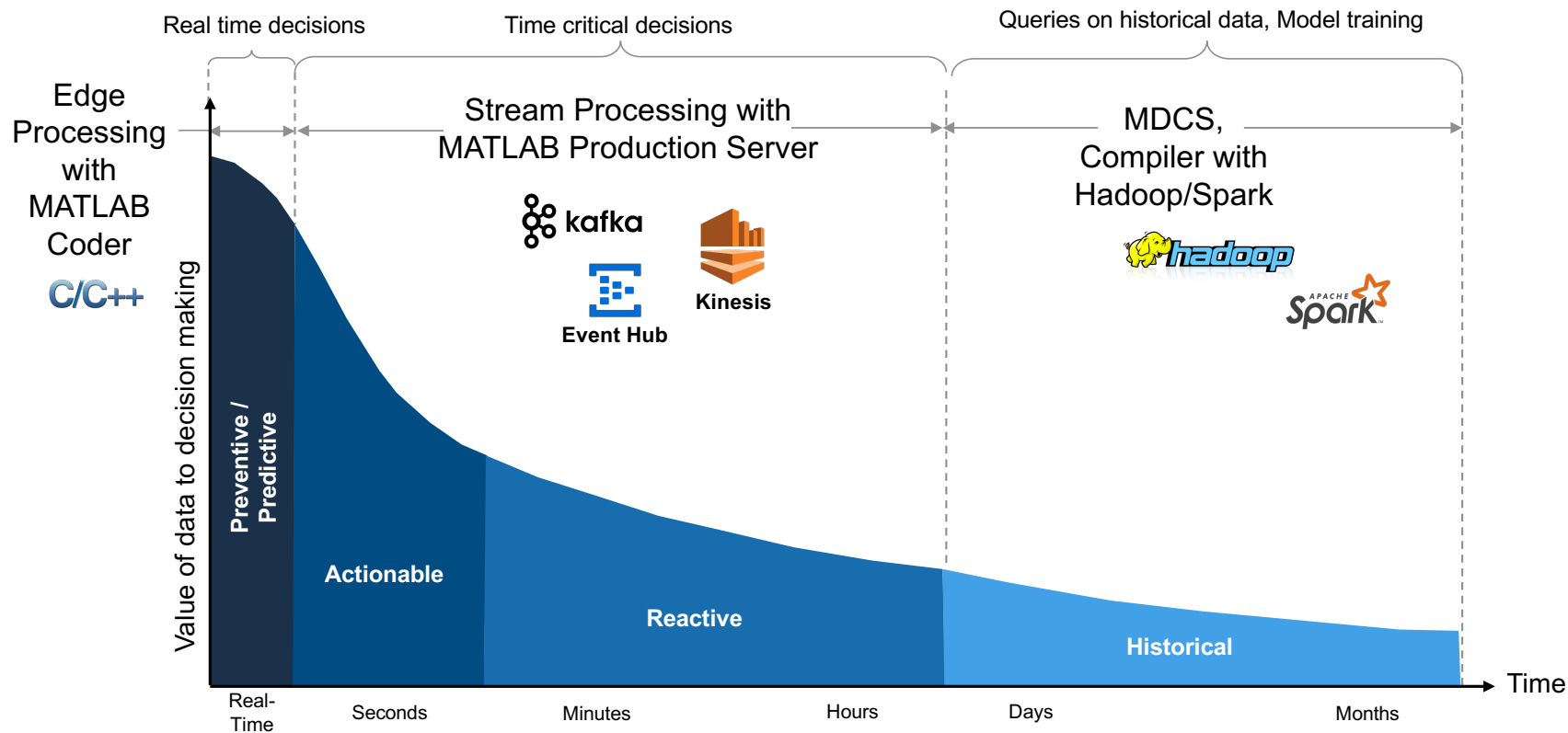
- **Batch Processing** applies computation to a finite sized historical data set that was acquired in the past



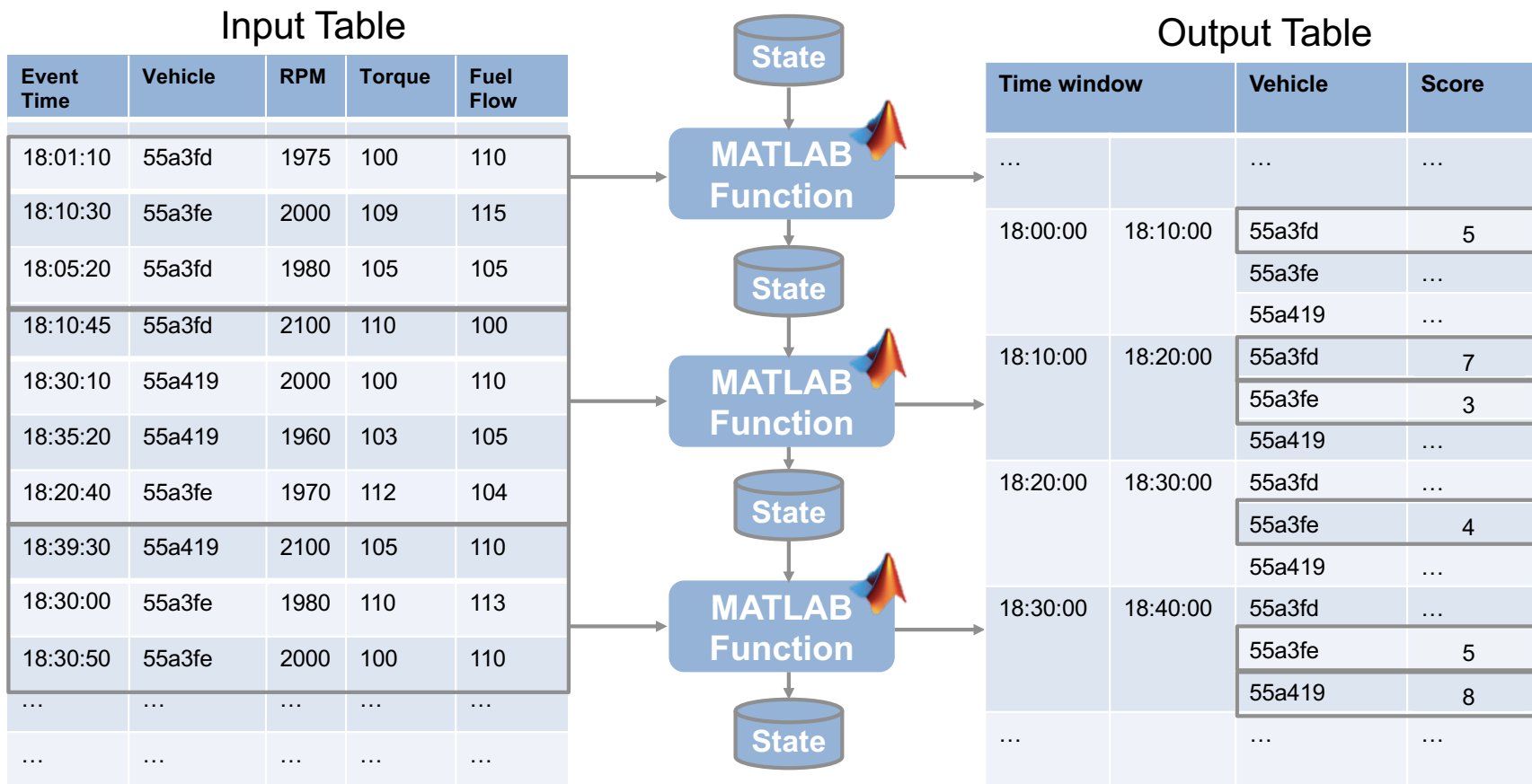
- **Stream Processing** applies computation to an unbounded data set that is produced continuously



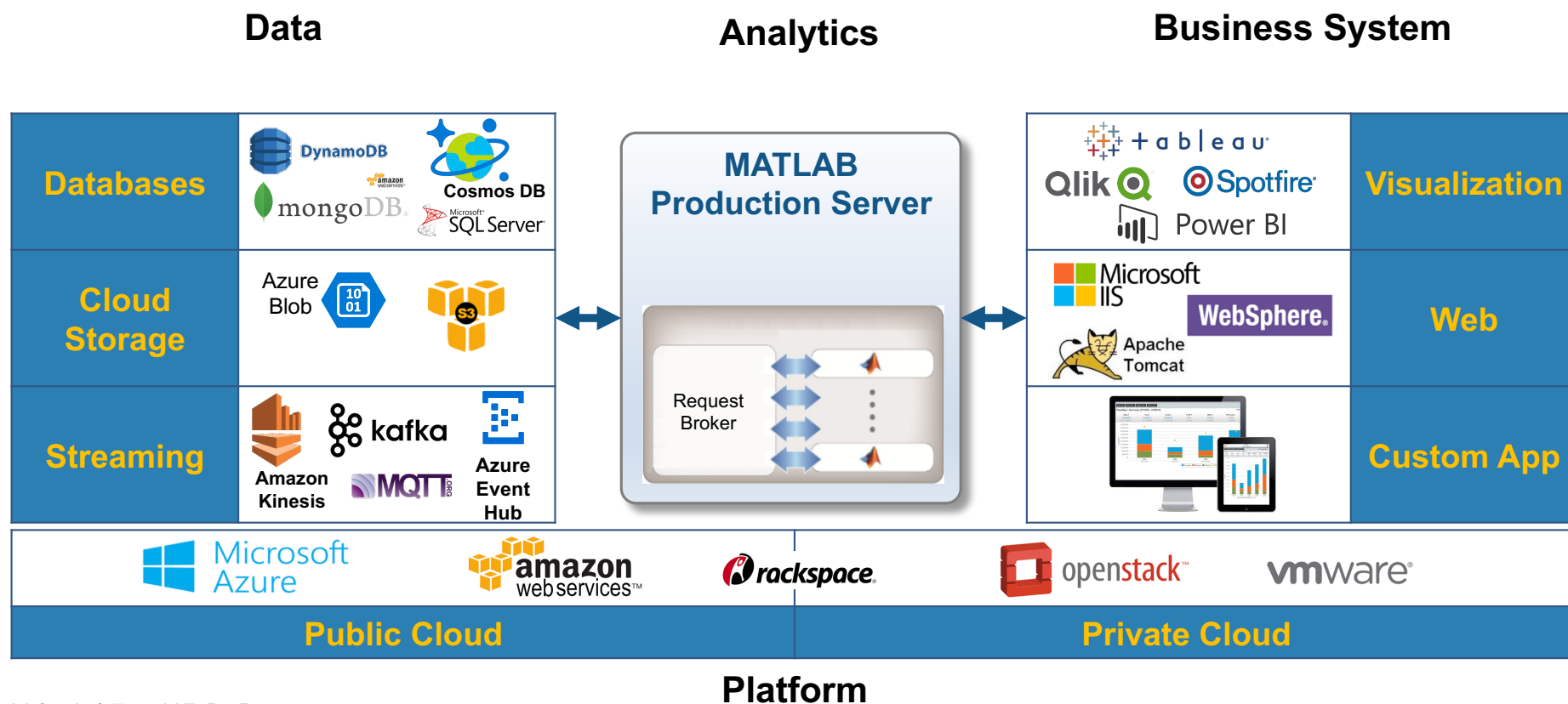
Stream processing exploits the fact that recent data tends to be more valuable



Streaming data is treated as an unbounded table



Introducing MATLAB Production Server



MATLAB EXPO 2017

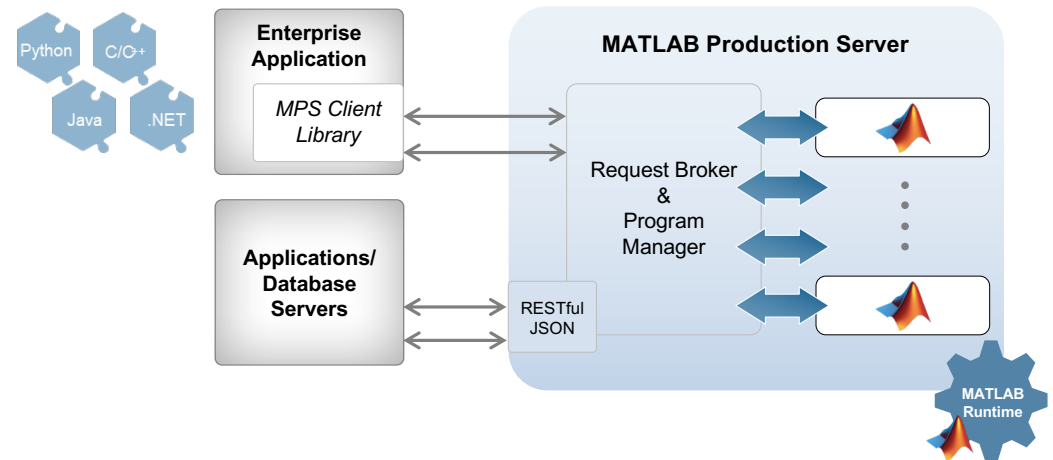
Introducing MATLAB Production Server

- Server software
 - Manages packaged MATLAB programs and worker pool

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

- RESTful JSON interface

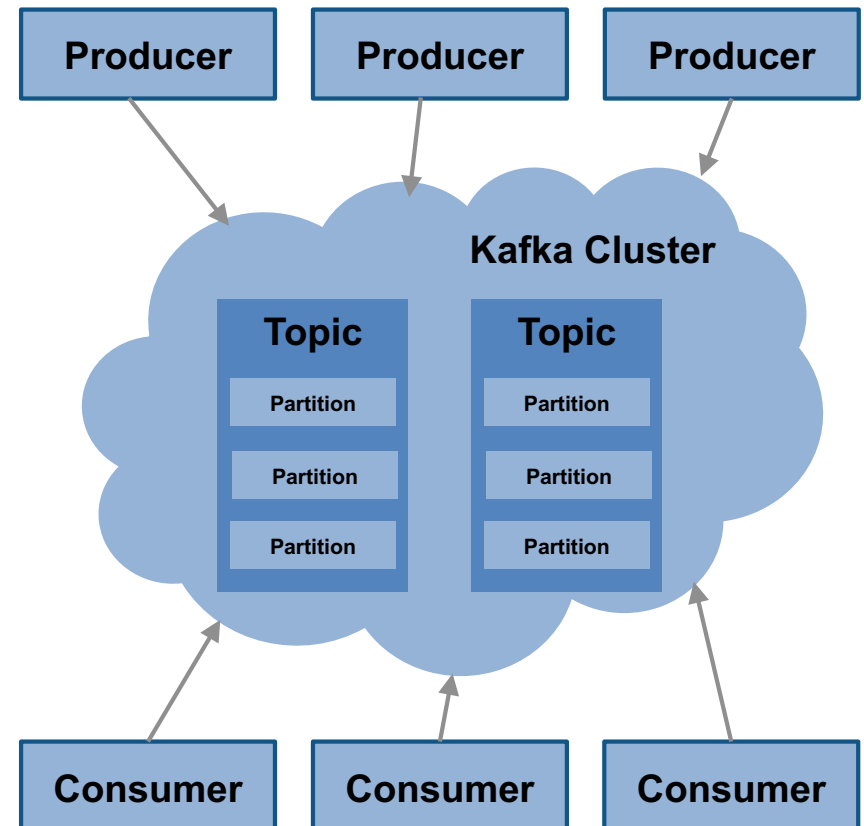
- Lightweight client libraries
 - C/C++, .NET, Python, and Java



Introducing Apache Kafka

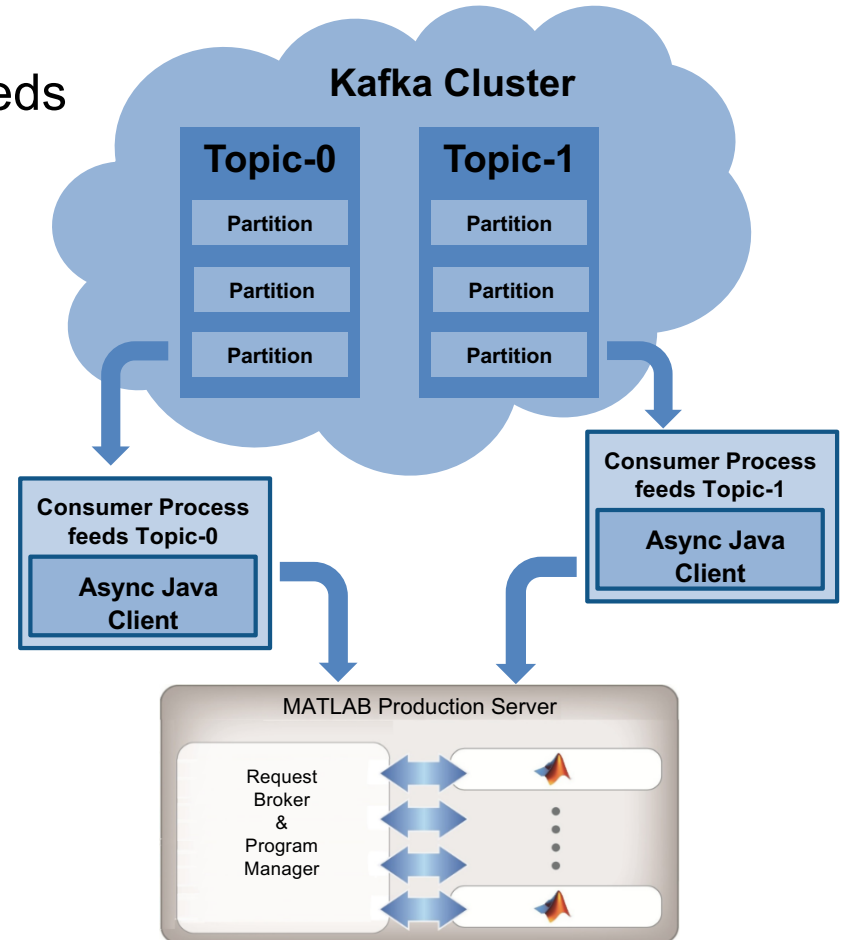
- Kafka is a high through-put distributed messaging system
- Originally developed at LinkedIn and open sourced in 2011
- Kafka is architected as a massively scalable publish/subscribe message queue
- Well suited for large scale streaming applications

MATLAB EXPO 2017

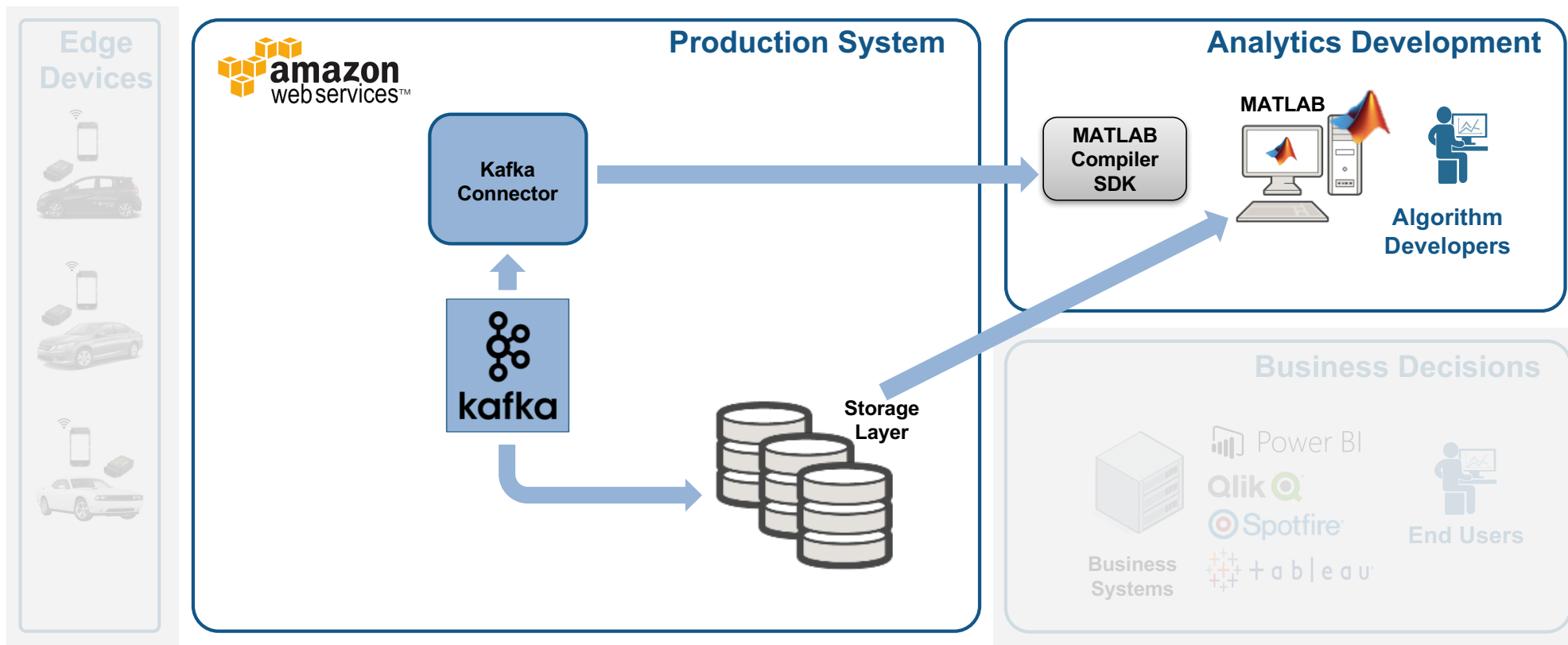


Connecting MATLAB Production Server to Apache Kafka

- 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!



Develop, Test, and Deploy a Stream Processing Function



Develop a Stream Processing Function in MATLAB

DEMO

Test Your Stream Processing Function on Live Data

The screenshot shows the MATLAB Compiler SDK interface for a project named 'kafkaconsumer.pj'. The 'TEST' tab is active, displaying the following components:

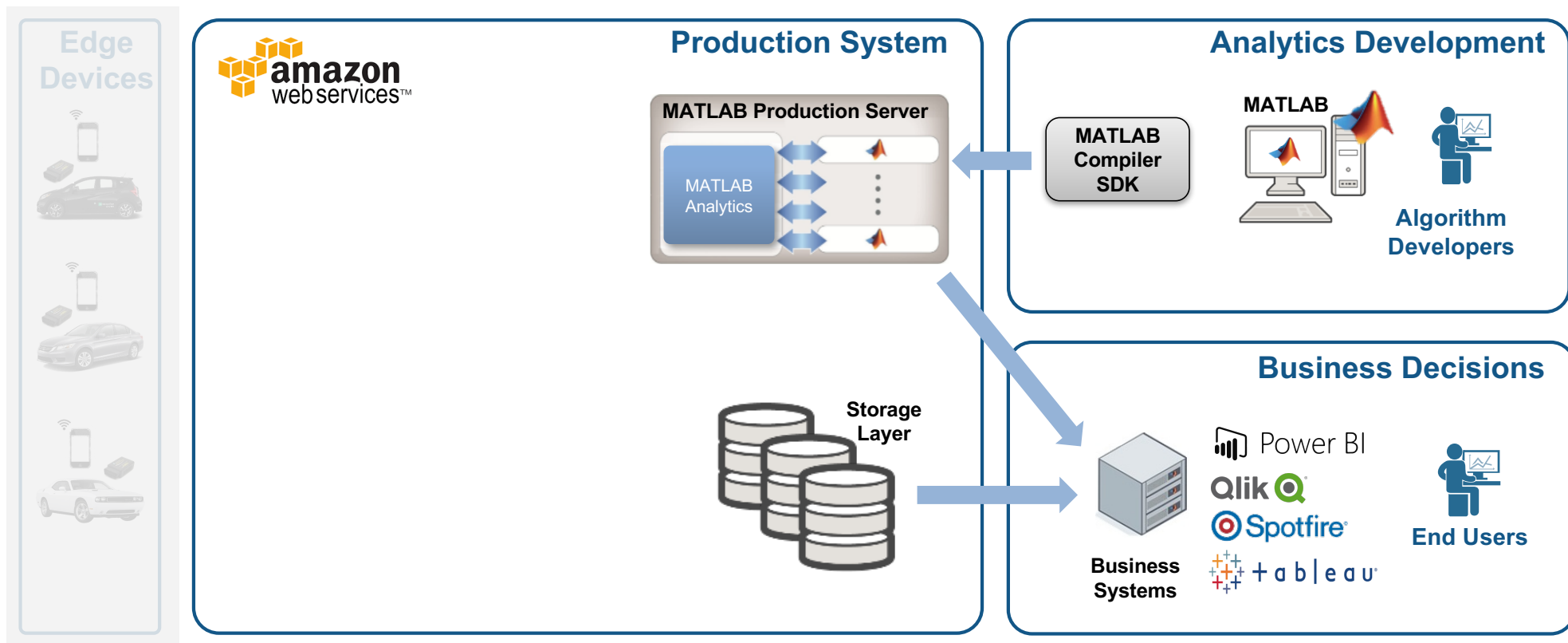
- Port:** 9910
- Enable CORS:**
- Server Address:** Accepting client connections on: http://localhost:9910/kafkaconsumer
- Server Requests:** A table showing the status of five requests.
- Server Log:** A log window showing system messages and timestamps.

ID	Function	Status
0	consume(msg)	Complete
1	consume(msg)	Complete
2	consume(msg)	Complete
3	consume(msg)	Complete
4	consume(msg)	Pending

```

83 [Tue Oct 31 16:56:18 EDT 2017] bound to [::]:9910
84 [Tue Oct 31 16:56:18 EDT 2017] [SERVICE:date_clock] Starting
85 [Tue Oct 31 16:56:18 EDT 2017] [SERVICE:date_clock] Started successfully
86 [Tue Oct 31 16:56:18 EDT 2017] [SERVICE:http] [endpoint:::9910] Starting
87 [Tue Oct 31 16:56:18 EDT 2017] [SERVICE:http] [endpoint:::9910] Waiting for a connection on port 9910
88 [Tue Oct 31 16:56:18 EDT 2017] [SERVICE:http] [endpoint:::9910] Started successfully
89 [Tue Oct 31 16:56:18 EDT 2017] [SERVICE:http] [endpoint:::9910] Listening for connections
90 [Tue Oct 31 16:58:15 EDT 2017] [SERVICE:http] [endpoint:::9910] Accepted connection from [::ffff:127.0.0.1]:55935 (socket = 15276)
91 [Tue Oct 31 16:58:15 EDT 2017] [SERVICE:http] [endpoint:::9910] Waiting for a connection on port 9910
    
```

Complete Your Application



Complete Your Application

The screenshot shows the MATLAB R2017b environment. The Live Editor displays the following code:

```

Read Fleet Data from Database

server = 'fleet.mwlab.io';
port = 27017;
dbname = 'fleet_production';
conn = mongo(server, port, dbname);

Fetch data

collection = "trips";
tripData = find(conn, collection);
    
```

The Command Window shows the output of the code:

```

>> tripData(1:5,:)
ans =
    x_id                created_at                k5                kc                kff1001            kff100
    _____
    '55a41cbc69702d115b05b508' 'Mon Jul 13 13:17:00 PDT 2015' []                3634                [113.1120]          -122.1
    '55a41cbc69702d115b05b509' 'Mon Jul 13 13:17:00 PDT 2015' [84]              3634                [113.0400]          -122.1
    '55a41cbc69702d115b05b50a' 'Mon Jul 13 13:17:00 PDT 2015' [84]              3780                [113.7960]          -122.1
    '55a41cbc69702d115b05b50b' 'Mon Jul 13 13:17:00 PDT 2015' [84]              3780                [114.7320]          -122.1
    '55a41cbc69702d115b05b50c' 'Mon Jul 13 13:17:00 PDT 2015' [84]              3780                [115.7760]          -122.1
    
```

The Web Map Display window shows a map of New York City with a blue line representing a trip route. The geographic coordinates are 40.70012, -73.68916.

Reference: <https://www.mathworks.com/help/database/ug/import-and-analyze-data-from-mongodb.html>

MATLAB EXPO 2017

Go Live!

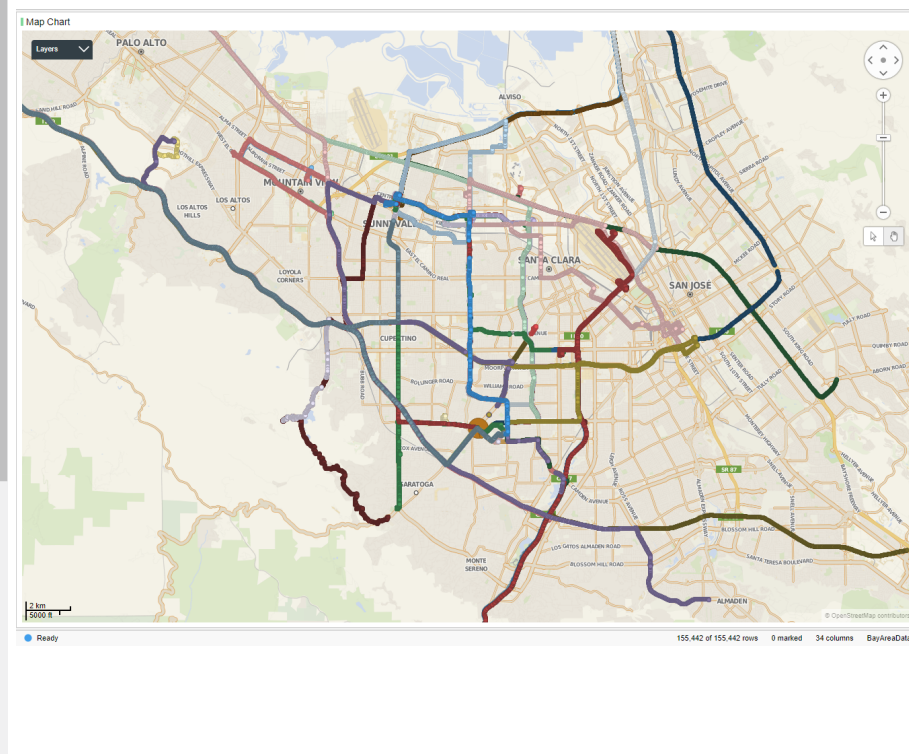
Fleet Summary

Automatic updates: Enabled

Fleet Statistics

Acceleration/Deceleration Events, 2014 - 2017

Category	Percentage
Aggressive deceleration	11.3%
Safe acceleration	9.9%
Moderate acceleration	5.2%
Aggressive acceleration	3.6%
Other	69.1%

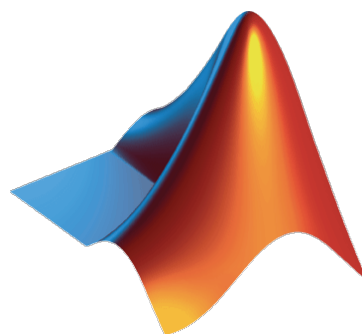


Key Takeaways

- 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](#)
 - <https://www.mathworks.com/help/database/ug/import-and-analyze-data-from-mongodb.html>



MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders."