

# Big Engineering Data Analytics with MATLAB

Dmitrij Martynenko

Application Engineer

[dmitrij.martynenko@mathworks.de](mailto:dmitrij.martynenko@mathworks.de)

MathWorks

**AUTOMOTIVE CONFERENCE 2015**



## How do you define Big Data?

*“Any collection of data sets so large and complex that it becomes difficult to process using ... traditional data processing applications.”*

(General Definition)

*“Any collection of data sets so large that it becomes **difficult** to process using traditional MATLAB functions, which assume all of the data is **in memory**.”*

(MATLAB)

# Your Big Data Sources

## Vehicle Component

Test or Simulation Method

	Engine Control Module (ECM)	Engine	Exhaust Gas Aftertreatment	Vehicle and Driver	Topographic Drive Route	Variability of Results
Fleet of Test Vehicles	Blue	Blue	Blue	Blue	Blue	
Vehicle in Loop	Blue	Blue	Blue	Blue	Yellow	
Aftertreatment in Loop	Blue	Blue	Blue	Yellow	Yellow	
Engine in Loop	Blue	Blue	Yellow	Yellow	Yellow	
ECM in Loop	Blue	Yellow	Yellow	Yellow	Yellow	
Pure Simulation	Yellow	Yellow	Yellow	Yellow	Yellow	

Hardware

Software

# Challenges Analyzing Fleet Data

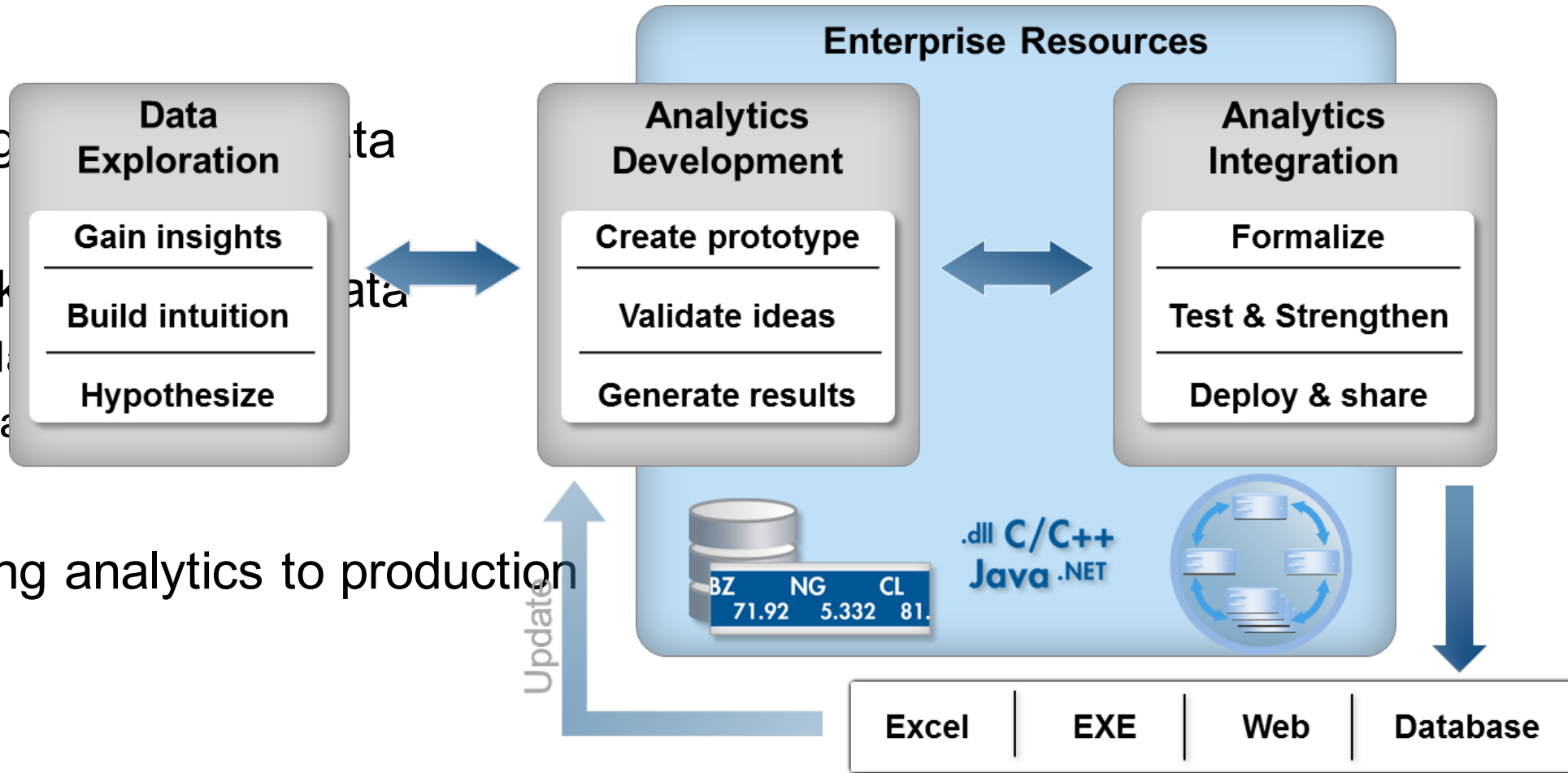
**More Data** ⇒ **Better Understanding of Field Conditions**  
**More Interesting Events**

**Challenges:**

- Big Data**
- Needle in the Haystack**
- Testing Ideas**
- Knowledge Transfer**

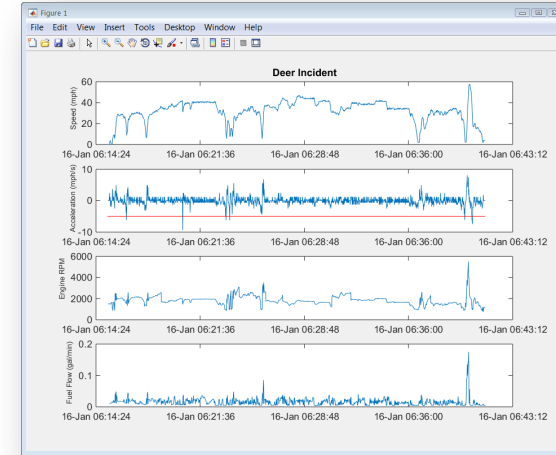
# Data Analytics with MATLAB

- Merge data
- Work with data
  - Merge
  - Handle
- Taking analytics to production



# Fleet Data Event Detection

- Parse data, find sudden deceleration
- MapReduce workflow



**Datastore**

Device Time	Engine RPM	Fuel flow rate	Torque	Speed (ODB)
15-Jan-XXXX 15:37:24	1445.75	0.0537	5.26	3.10
15-Jan-XXXX 15:37:25	1445.75	0.0537	5.74	0.48
15-Jan-XXXX 15:37:26	1445.75	0.0537	17.80	4.97
15-Jan-XXXX 15:37:27	1458.5	0.0107	11.10	6.21
15-Jan-XXXX 15:37:28	1458.5	0.0107	11.10	0.59

**Map**

Acceleration	Decelaration
-2.62	0
4.49	0
1.24	0
-5.62	1

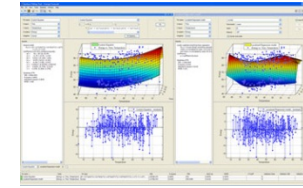
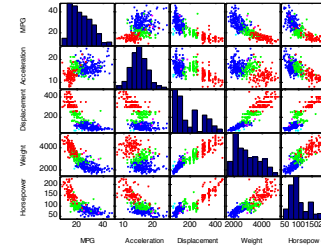
**Reduce**

Acceleration
-5.62

# Analysis Domains

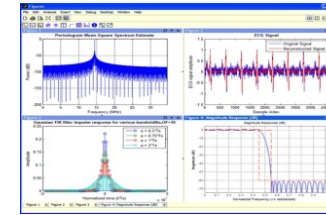
## Statistics

- Summary Statistics
- Regression, ANOVA, Machine Learning



## Signal Processing

- Sound quality analysis
- LIDAR analysis



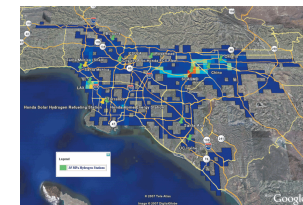
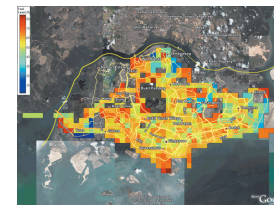
## Image Processing

- Active Safety

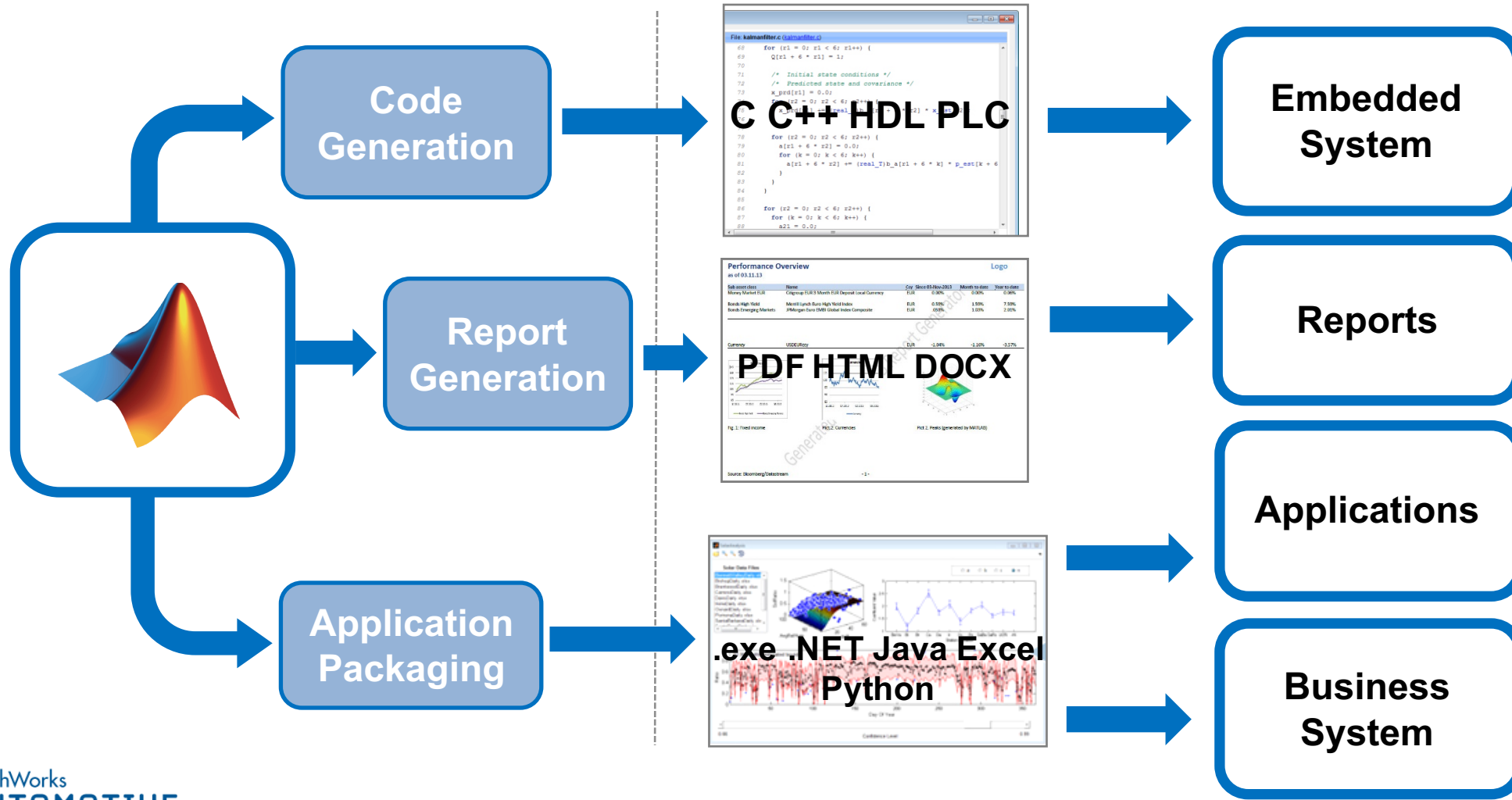


## Location/Mapping

- Analyzing GPS Data
- Custom Visualizations



# Taking MATLAB to Production





# Analyzing Big Data using MATLAB

- Operate on Big Data from MATLAB with MapReduce
- Quickly try out an idea, then iterate
- Transfer your results by taking MATLAB to production



# Additional Resources

## Machine Learning

**Machine Learning with MATLAB**

Build predictive models and discover useful patterns from observed data.

[Watch video](#)

Machine learning algorithms use computational methods to "learn" information directly from data without assuming a predetermined equation as a model. They can adaptively improve their performance as you increase the number of samples available for learning.

Machine learning algorithms are used in applications such as computational finance (credit scoring and algorithmic trading), computational biology (tumor detection, drug discovery, and DNA sequencing), energy production (price and load forecasting), natural language processing, speech and image recognition, and advertising and recommendation systems.

Machine learning is often used in big data applications, which have large datasets with many predictors (features) and are too complex for a simple parametric model. Examples of big data applications include forecasting electricity load with a neural network, or bond rating classification for credit risk using an ensemble of decision trees.

**Classification**  
Build models to classify data into different categories.

**Regression**  
Build models to predict continuous data.

**Clustering**  
Find natural groupings and patterns in data.

[mathworks.com/machine-learning](http://mathworks.com/machine-learning)

## Parallel Computing

**Parallel Computing**

Perform large-scale computations using multicore desktops, GPUs, clusters, grids, and clouds.

Explore Products for Parallel Computing  
MATLAB Distributed Computing Server™  
Parallel Computing Toolbox™

Parallel Computing Resources  
Webinars  
Training  
Technical Articles

Large-scale simulations and data processing tasks that support engineering and scientific activities such as mathematical modeling, algorithm development, and testing can take an unreasonably long time to complete or require a lot of computer memory.

You can speed up these tasks by taking advantage of high-performance computing resources, such as multicore computers, GPUs, computer clusters, and grid and cloud computing services.

MathWorks parallel computing products let you use these resources from MATLAB® and Simulink® without making major changes to your computing environment and workflows. Using parallel computing products, you can:

- Reduce your programming effort
- Run an application across a range of high-performance computing resources
- Program and execute parallel code interactively or in batch mode

Listening to the World's Oceans: Searching for Marine Mammals by Detecting and Classifying Terabytes of Bioacoustic Data in Clouds of Noise

<http://www.mathworks.com/solutions/parallel-computing/>

## MapReduce

**MATLAB MapReduce and Hadoop**

Tackling Big Data with MATLAB

[Watch video](#)

MATLAB® has numerous capabilities for exploring and analyzing big data sets. Among them is MapReduce, a powerful, and established programming technique for applying filtering, statistics and other general analysis methods to big data.

The MapReduce functionality built into MATLAB lets you analyze data that does not fit into memory. By running your MapReduce based algorithms in parallel (using Parallel Computing Toolbox™), you can better utilize the processing resources on your desktop without changing your algorithms.

To analyze data in MATLAB using MapReduce:

1. Specify the data you want to analyze using datastore
2. Create your map and reduce functions in MATLAB
3. Execute your map and reduce functions using mapreduce

While MATLAB MapReduce is optimized for array-based analysis, it is fully compatible with Hadoop MapReduce, so you can run your MapReduce based algorithms within the Hadoop MapReduce framework.

- Execute MapReduce based algorithms on Hadoop directly from the MATLAB desktop, using MATLAB Distributed Computing Server™
- Package MapReduce based algorithms for deploying to production Hadoop systems, using MATLAB Compiler™

<http://www.mathworks.com/discovery/matlab-mapreduce-hadoop.html>

# Questions?