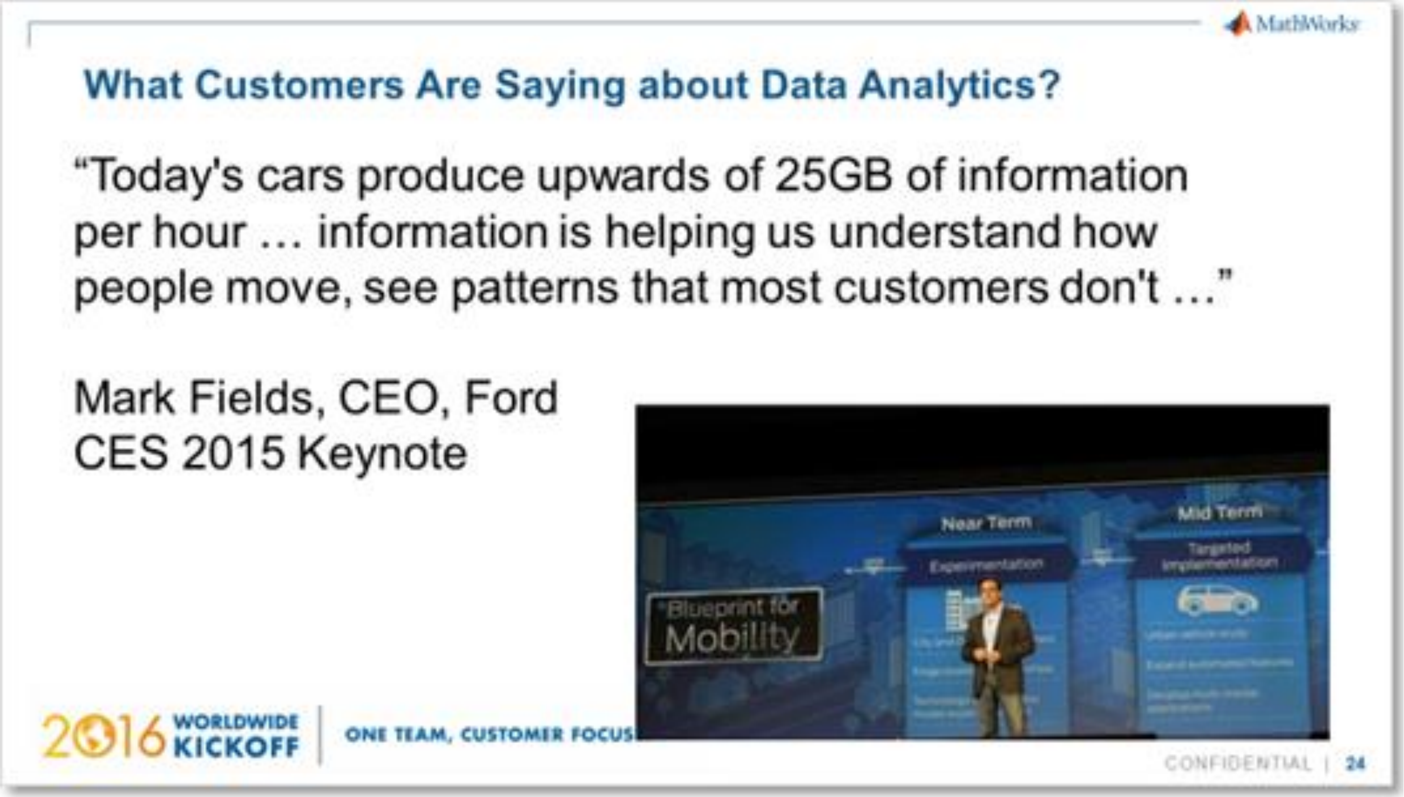
TeraSoft 鈦思科技 The Power of Driver for the Best!!

Big Data in MATLAB

Application Engineer
Jeffrey Liu



MathWorks

What Customers Are Saying about Data Analytics?

“Today's cars produce upwards of 25GB of information per hour … information is helping us understand how people move, see patterns that most customers don't …”

Mark Fields, CEO, Ford
CES 2015 Keynote

Blueprint for
Mobility
Near Term
Experimentation
Mid Term
Targeted implementation

2016 WORLDWIDE KICKOFF | ONE TEAM, CUSTOMER FOCUS

CONFIDENTIAL | 24

Why doing data analysis?

More Data ⇨

Better Understanding of Field Conditions

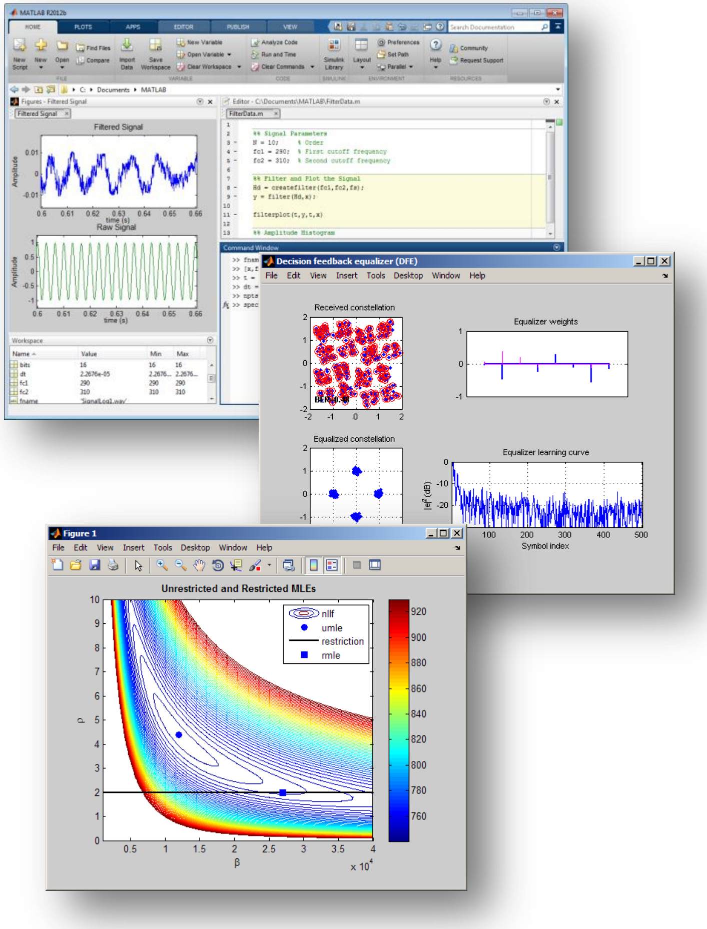
More Interesting Events

Challenges:

Big Data
Needle in the Haystack
Testing Ideas
Knowledge Transfer

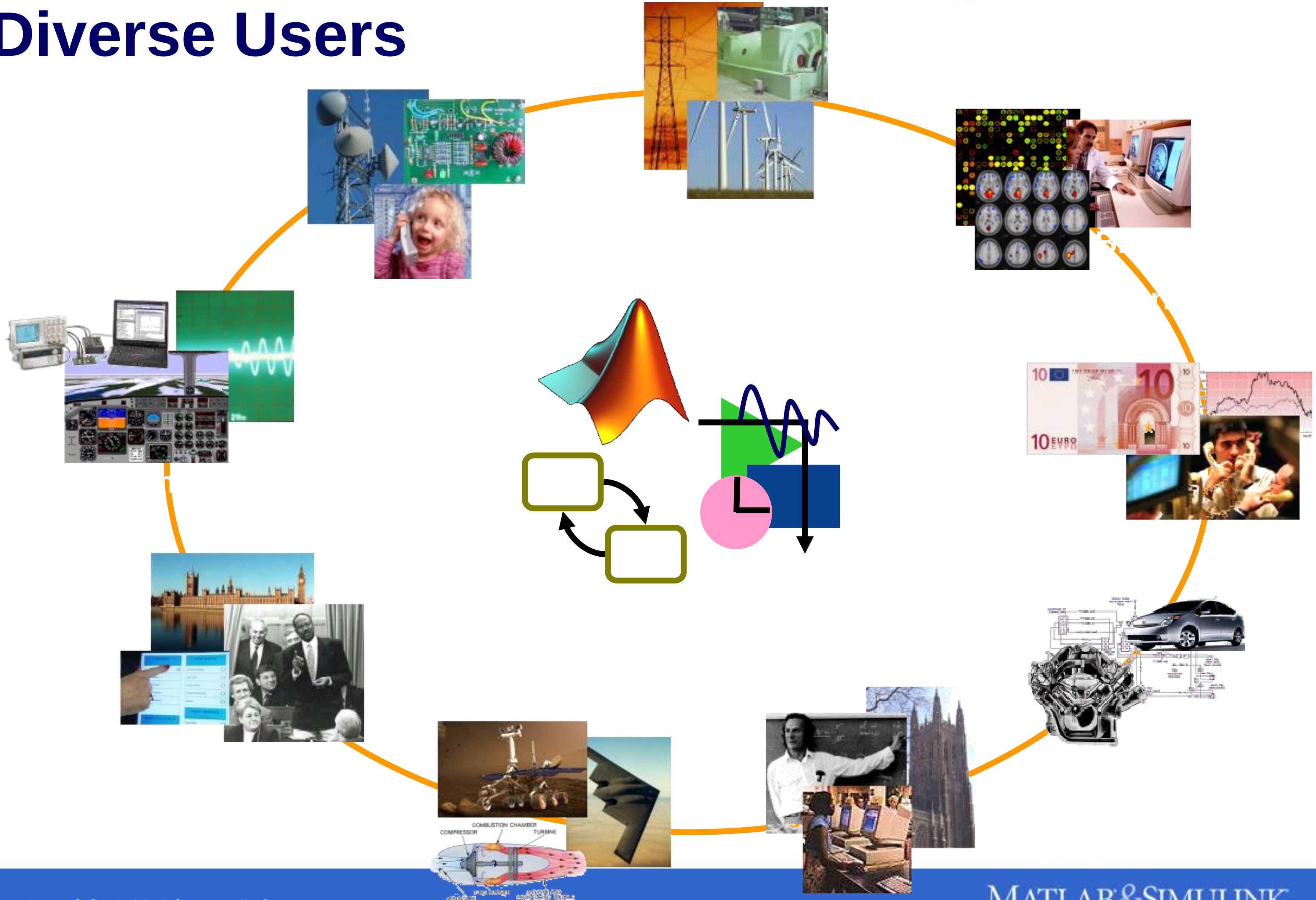
What is MATLAB?

- High-level language
- Interactive development environment
- Used for:
  - Numerical computation
  - Data analysis and visualization
  - Algorithm development and programming
  - Application development and deployment

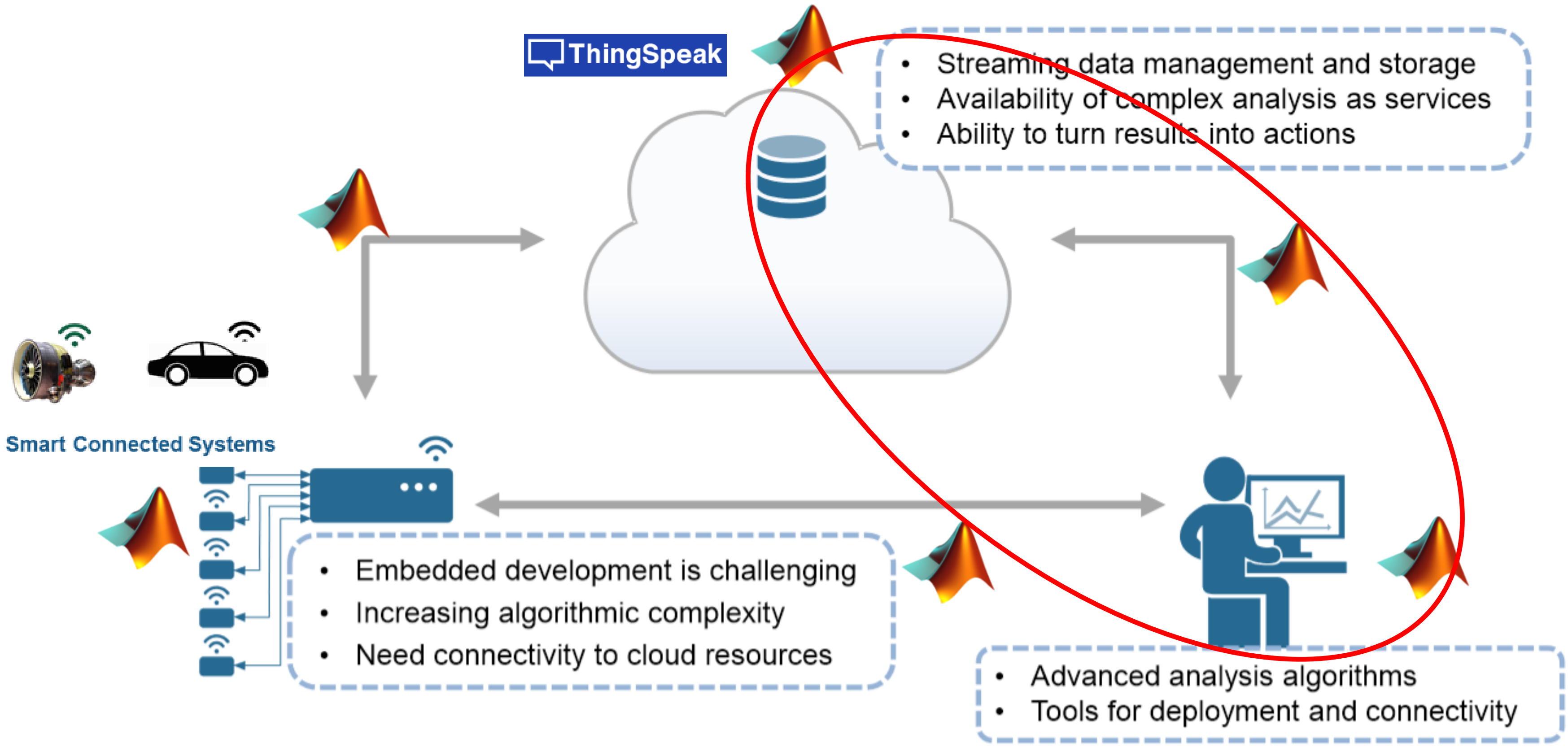


```
%% Signal Parameters
N = 10;      % Order
fc1 = 290;   % First cutoff frequency
fc2 = 310;   % Second cutoff frequency

%% Filter and Plot the Signal
Hd = createfilter(fc1,fc2,fs);
y = filter(Hd,x);

filterplot(t,y,t,x)

%% Amplitude Histogram
```

Diverse Users



Internet of Things and Big Data



- Streaming data management and storage
- Availability of complex analysis as services
- Ability to turn results into actions

- Embedded development is challenging
- Increasing algorithmic complexity
- Need connectivity to cloud resources

- Advanced analysis algorithms
- Tools for deployment and connectivity

Big Data in Industry

ENERGY
Asset Optimization

FINANCE
Market Risk, Regulatory

AUTO
Fleet Data Analysis

AERO
Maintenance, reliability

Medical Devices
Patient Outcomes

Challenges of Big Data

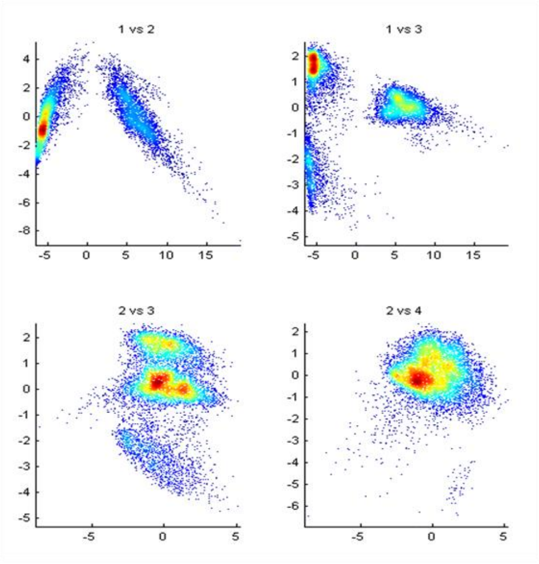
“Any collection of data sets so large and complex that it becomes difficult to process using … traditional data processing applications.”
(Wikipedia)

- How to get started
- Rapid data exploration
- Development of scalable algorithms
- Use of algorithms within business systems

New Big Data Capabilities in MATLAB

Memory and Data Access

- 64-bit processors
- Memory Mapped Variables
- Disk Variables
- Databases
- **Datastores** R2014b



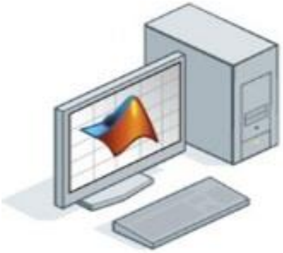
Programming Constructs

- Streaming
- Block Processing
- Parallel-for loops
- GPU Arrays
- SPMD and Distributed Arrays
- **MapReduce** R2014b

Platforms

- Desktop (Multicore, GPU)
- Clusters
- Cloud Computing (MDCS on EC2)
- **Hadoop** R2014b

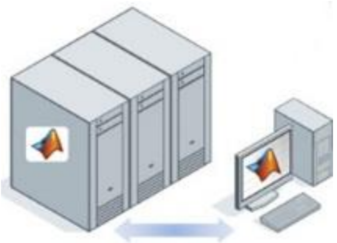
Big Data on the Desktop



- Expand workspace
  - *64 bit processor support – increased in-memory data set handling*
- Access portions of data too big to fit into memory
  - *Memory mapped variables – huge binary file*
  - *Datastore – huge text file or collections of text files*
  - *Database – query portion of a big database table*
- Variety of programming constructs
  - *System Objects – analyze streaming data*
  - *MapReduce – process text files that won’t fit into memory*
- Increase analysis speed
  - *Parallel for loops – use with multicore/multi-process machines*
  - *GPU Arrays*

Further Scaling Big Data Capacity

MATLAB supports a number of programming constructs for use with clusters



- General compute clusters
  - *Parallel for loops – embarrassingly parallel algorithms*
  - *SPMD – distributed processing*
- Hadoop clusters
  - *MapReduce – analyze data stored in the Hadoop Distributed File System*

Customer Example: Cornell Bioacoustics Scientists

- Challenge
  - Years of raw acoustic data—up to 10TB—recorded on multiple channels. This data must be filtered and processed to identify appropriate tolerances, compute time-optimized tracks for robots or flexible transfer systems
- Solution
  - Develop a high-performance computing platform for acoustic data analysis using MATLAB, Parallel Computing Toolbox, and MATLAB Distributed Computing Server
- Results
  - Years of development time saved
  - Analysis time reduced from weeks to hours
  - Previously unprocessed data analyzed in days

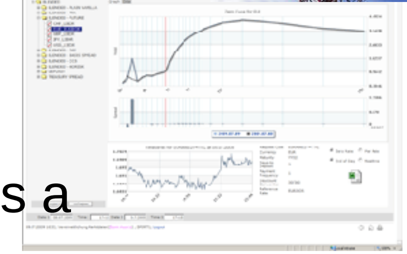
Case Study: Shell

- Challenges
  - 46 petabytes of data
  - Develop predictive analytics to gain actionable insights from big data
  - Real-time interventions when abnormalities are detected
- Solutions
  - MATLAB® Multivariate statistical models running on MATLAB Production Server™
- Result
  - Real-time batch and process monitoring
  - 6-8% uplift in production

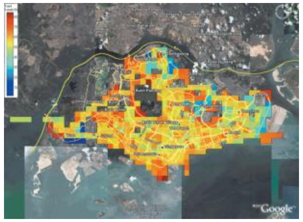
Other cases

- STIWA Increases Total Production Output of Automation Machinery

- Cognizant Speeds Customer Churn Analysis for Telecom Service Provider



- UniCredit Bank Austria Develops and Rapidly Deploys a Consistent, Enterprise-Wide Market Data Engine



- Analyzing Test Data from a Worldwide Fleet of Fuel Cell Vehicles at Daimler AG

- Edwards Air Force Base Accelerates Flight Test Data Analysis Using MATLAB and Parallel Computing

Strengths of MATLAB for Big Data analysis

Challenge	MATLAB Solution
Getting started	**Easy access to data from your desktop** Tools for accessing typical big data sets consisting of text or binary files, contained in database tables or stored on Hadoop
Rapid data exploration	**All the tools to explore and visualize data** Use all the power of MATLAB to explore and understand your data
Development of scalable algorithms	**Work on the desktop and scale to clusters** Tools for use in analyzing big data on your desktop, which scale for use on clusters, including Hadoop, if needed
Use within business systems	**Ease of deployment and leveraging enterprise** Push-button deployment into production including support for Hadoop