MATLAB for Data Analytics

MathWorks

Railway Systems

Automotive

Aeronautics

Retail

Finance

Off-highway vehicles

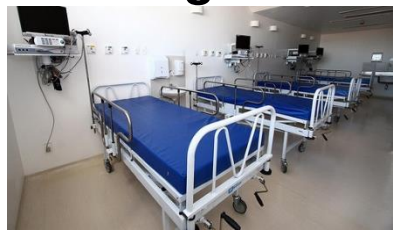
Industrial Automation

Oil & Gas

Clean Energy

Medical Devices

Healthcare Management



Logistics

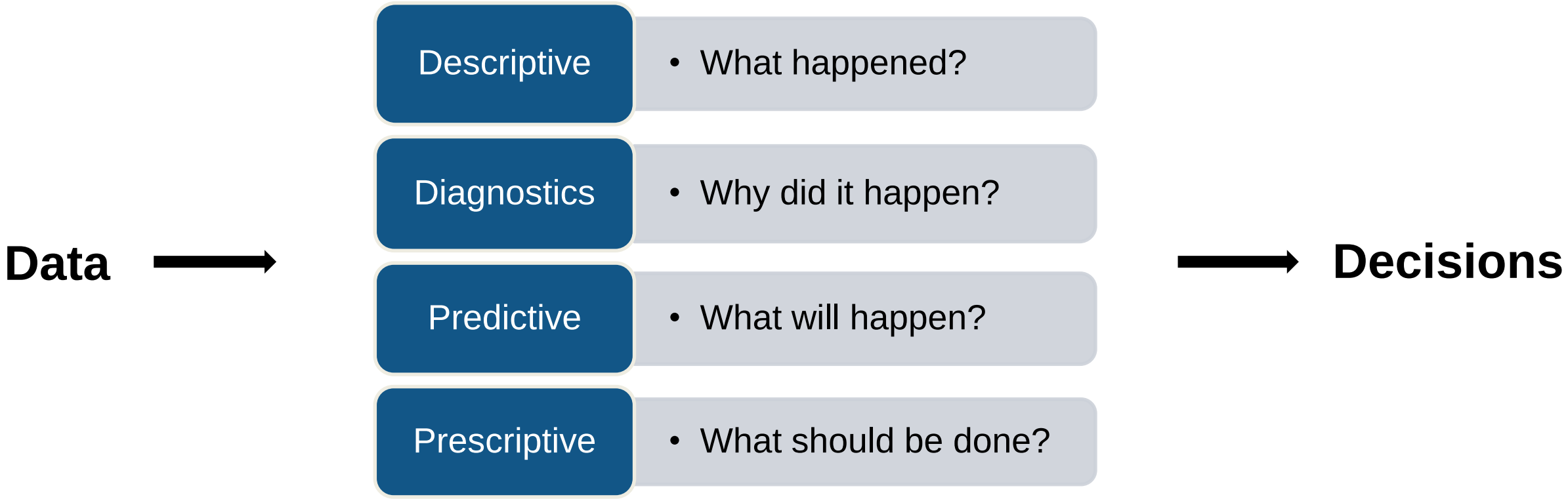
Internet

Condition Monitoring
Prognostics
Retail Analytics
Fleet Analytics
Operational Analytics
Process Analytics
Risk Analysis
Health Monitoring
Supply Chain
Mfg Process Analytics
Asset Analytics
Healthcare Analytics

Data Analytics

What is Data Analytics?

Turn large volumes of complex data into actionable information



Data Analytics Workflow

Access and Explore Data

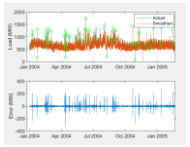
Files

Databases

Sensors

Preprocess Data

Working with Messy Data



Data Reduction/ Transformation

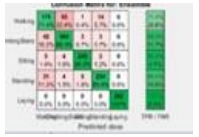
Feature Extraction

Develop Predictive Models

Model Creation e.g. Machine Learning

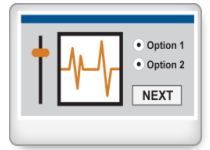
Parameter Optimization

Model Validation



Integrate Analytics with Systems

Desktop Apps



Enterprise Scale Systems

MATLAB Excel .NET .exe C/C++ Java .dll

Embedded Devices and Hardware

Data Analytics Workflow

Access and Explore Data

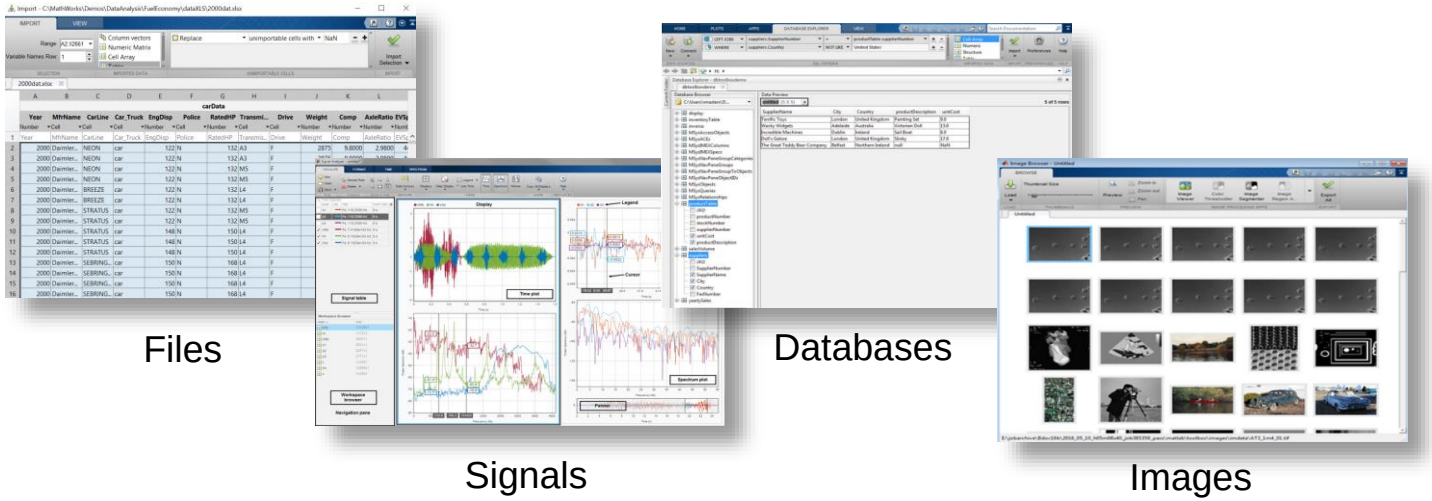
MATLAB Analytics work with business and engineering data 1

Preprocess Data

Files

Databases

Sensors

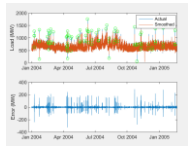


Files Signals Databases Images

- Point and click tools to access variety of data sources
- High-performance environment for big data

- Built-in algorithms for data preprocessing including sensor, image, audio, video and other real-time data

Working with Messy Data



Data Reduction/ Transformation

Feature Extraction

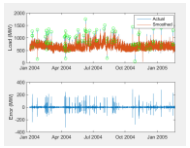
Data Analytics Workflow

Preprocess Data

MATLAB enables domain experts to do Data Science (2)

Develop Predictive Models

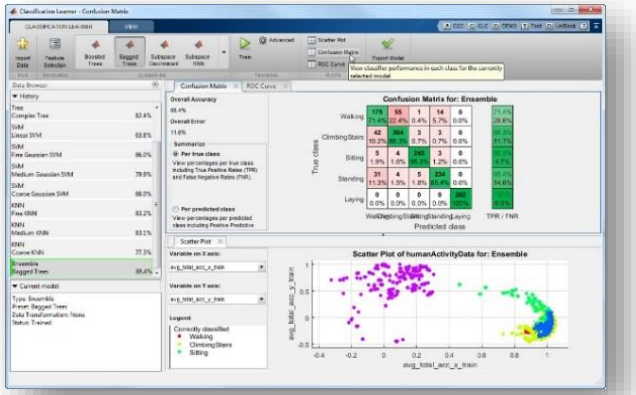
Working with Messy Data



Data Reduction/ Transformation

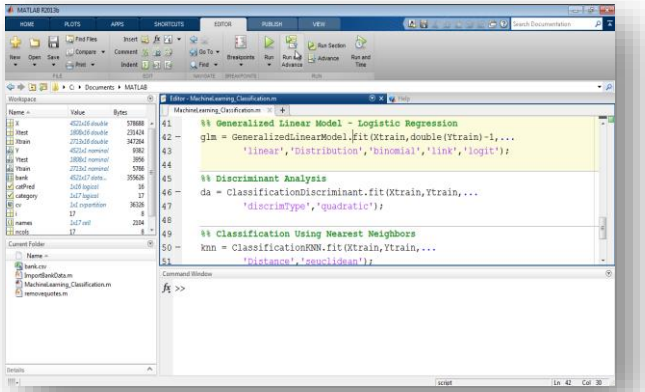
Feature Extraction

Apps



- Easy to use apps
- Wide breadth of tools to facilitate domain specific analysis
- Examples/videos to get started

Language

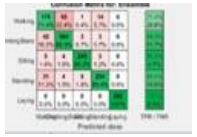


- Automatic MATLAB code generation
- High speed processing of large data sets

Model Creation e.g. Machine Learning

Parameter Optimization

Model Validation



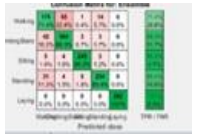
Data Analytics Workflow

Develop Predictive Models

Model Creation e.g. Machine Learning

Parameter Optimization

Model Validation

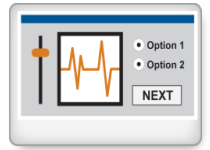


Challenges

- End user: Operators, Analysts, Administrative Staff, customers etc.
- Different target platforms:
  - Cluster or Cloud environment
  - Standalone desktop applications
  - Server based Web and enterprise systems
  - Embedded hardware
- Different Interfaces: C++, Java, Python, .NET etc.
- Need to translate analytics to production environment

Integrate Analytics with Systems

Desktop Apps



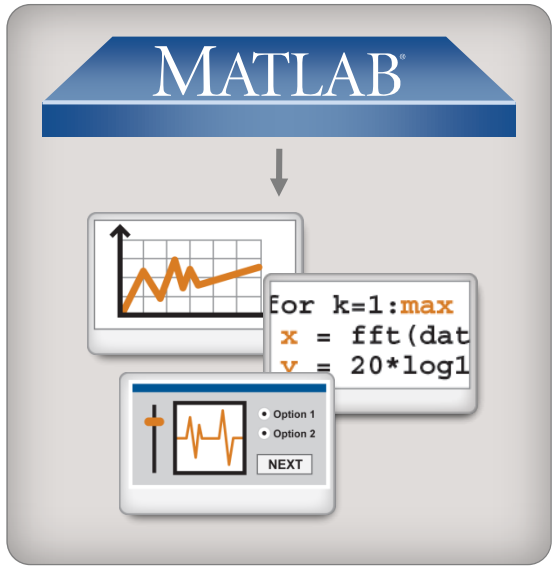
Enterprise Scale Systems

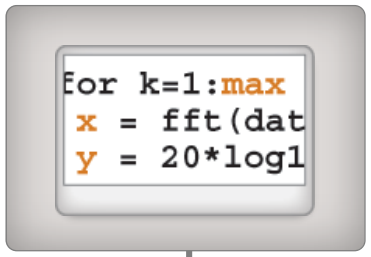
MATLAB Excel .NET .exe C/C++ Java .dll

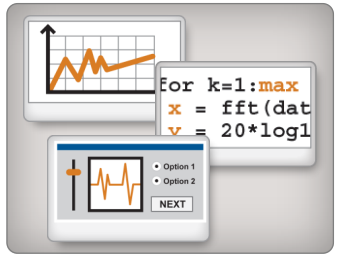
Embedded Devices and Hardware

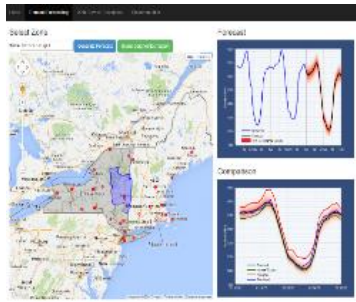
Integrate analytics with systems

MATLAB Analytics
run anywhere
3







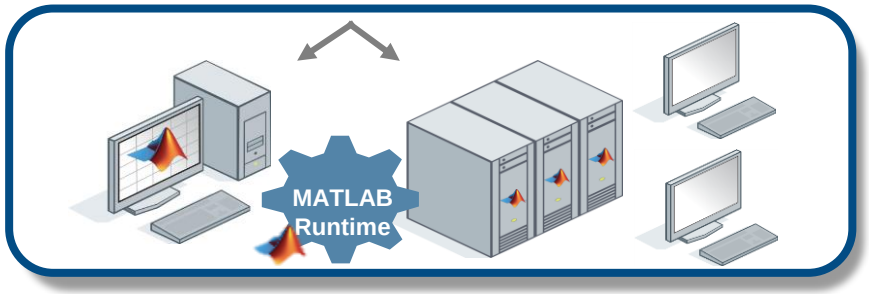


Embedded Hardware

C, C++ | HDL | PLC

Enterprise Systems

Standalone Application | Excel Add-in | Hadoop/ Spark | C/C++ | Java | Python | .NET | MATLAB Production Server

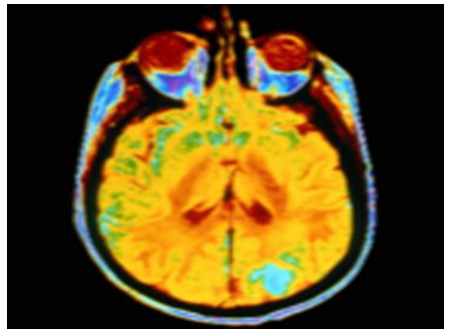


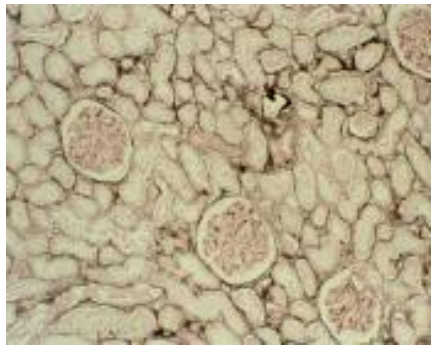
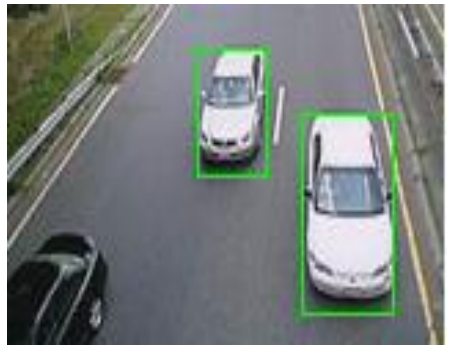
Key Takeaways

1. MATLAB Analytics work with **business and engineering data**
2. MATLAB enables **domain experts to do Data Science**
3. MATLAB Analytics **run anywhere**

Machine Learning is Everywhere

- Image Recognition
- Speech Recognition
- Stock Prediction
- Medical Diagnosis
- Data Analytics
- Robotics
- and more…

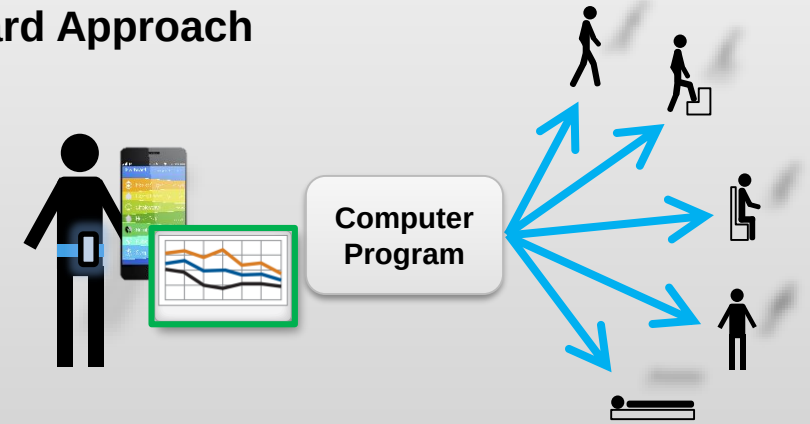


Machine Learning

Machine learning uses **data** and produces a **program** to perform a **task**

Task: Human Activity Detection

Standard Approach



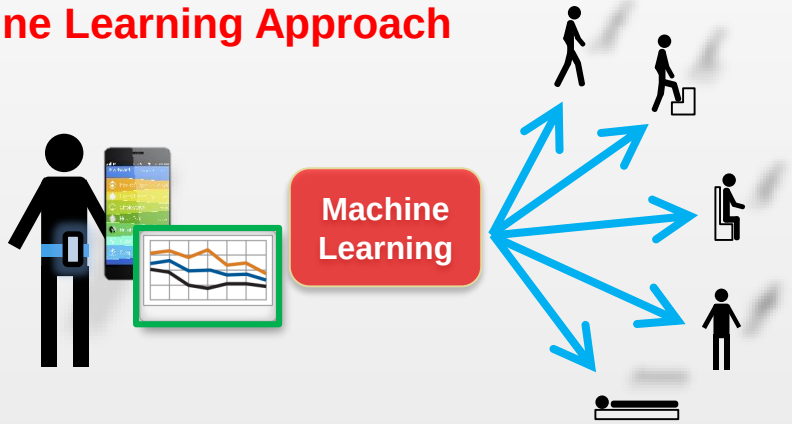
Hand Written Program

If X_acc > 0.5
then “SITTING”
If Y_acc < 4 and Z_acc > 5
then “STANDING”
…

Formula or Equation

$$Y_{activity} = \beta_1 X_{acc} + \beta_2 Y_{acc} + \beta_3 Z_{acc} + \dots$$

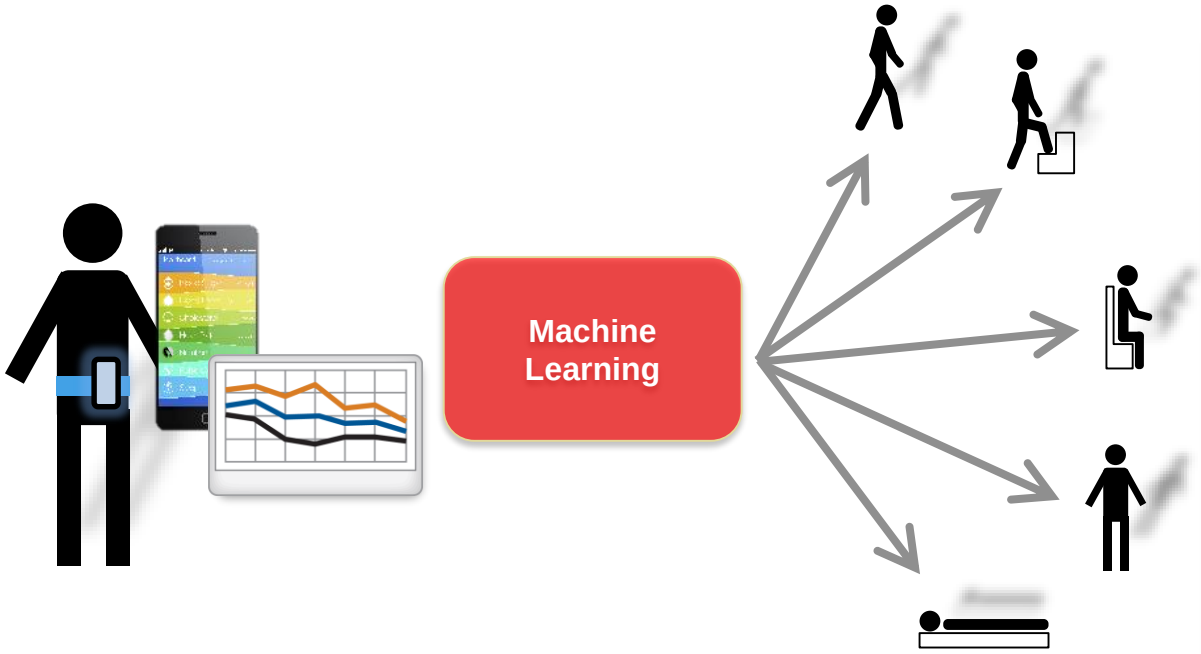
Machine Learning Approach

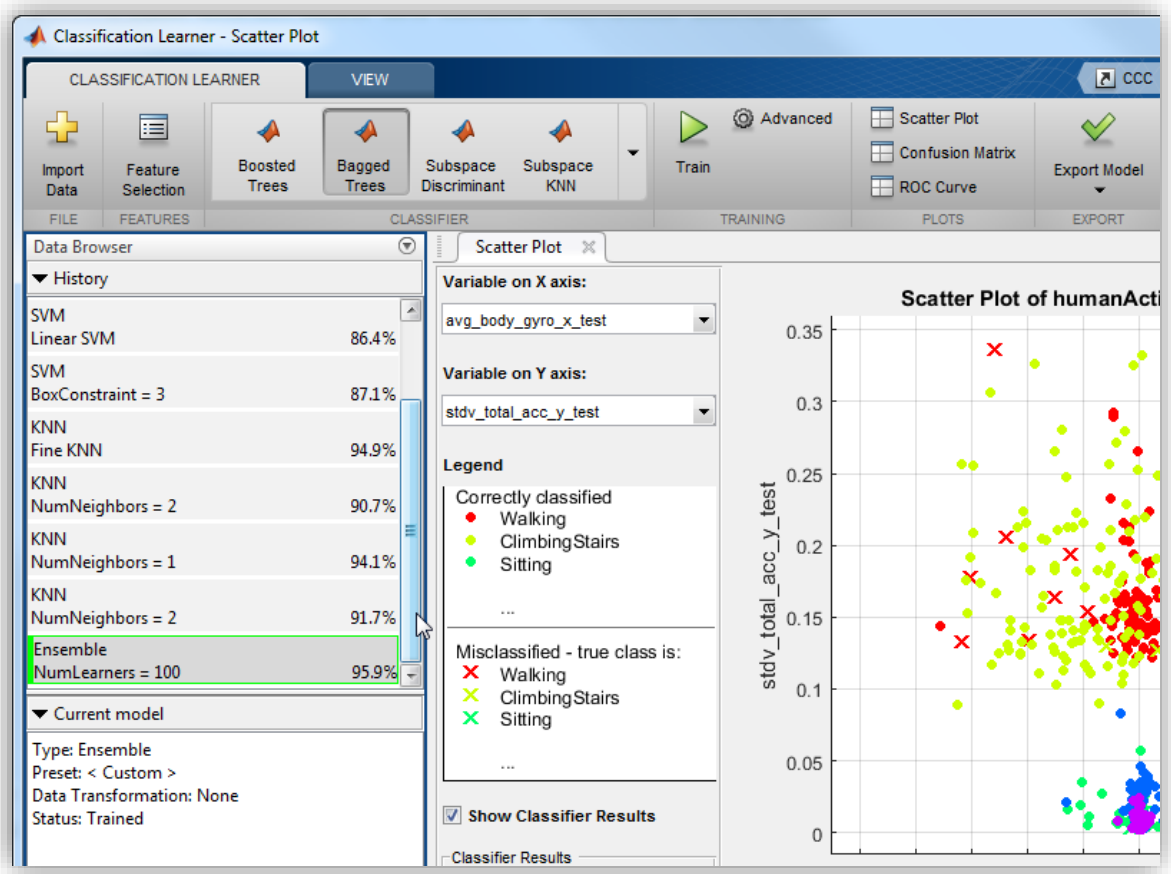


$model$: Inputs → Outputs

$$model = <^{\textbf{Machine Learning Algorithm}}>(sensor\_data, activity)$$

Example: Human Activity Learning Using Mobile Phone Data





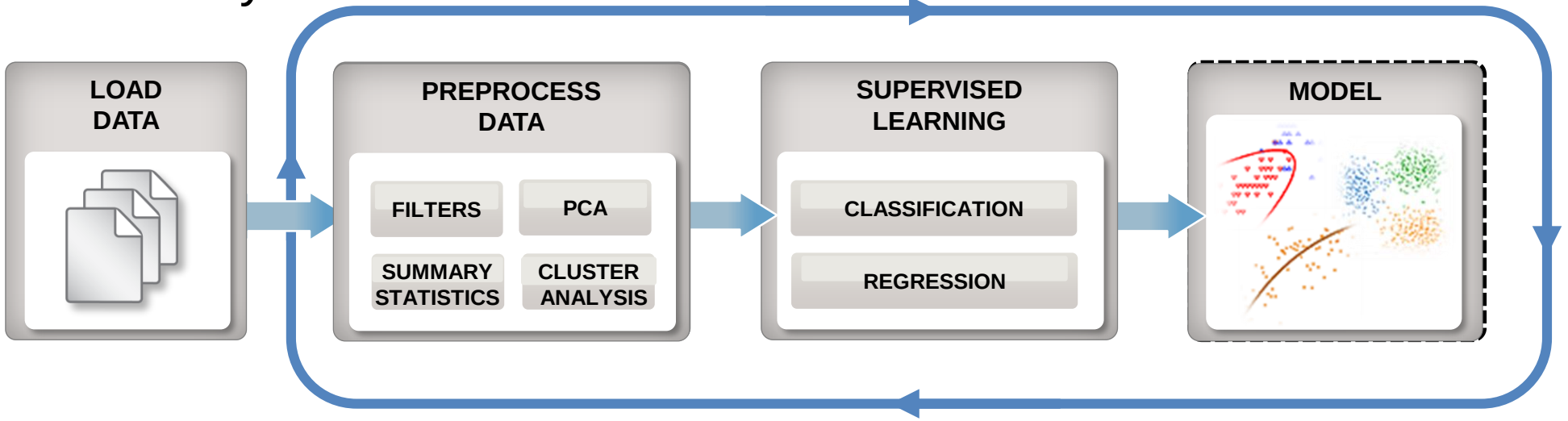
Data:

- 3-axial Accelerometer data
- 3-axial Gyroscope data

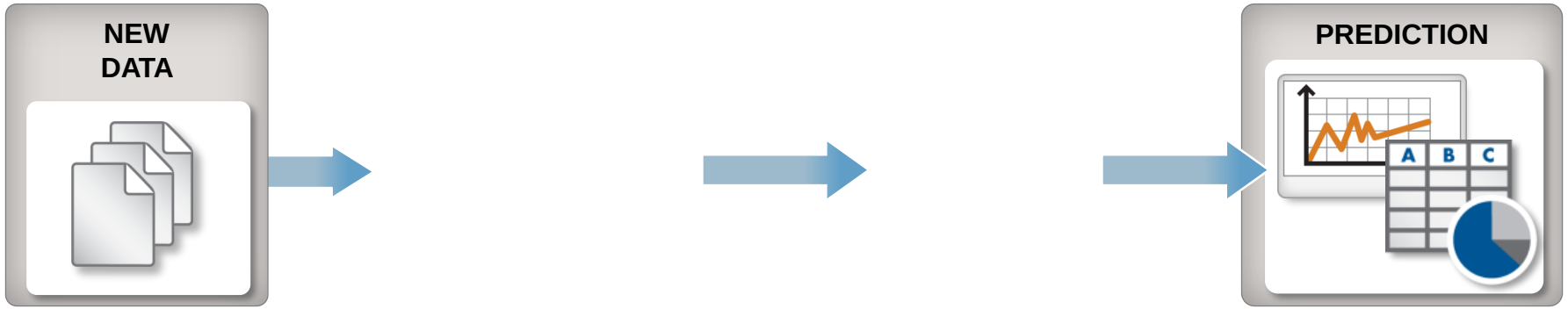
“essentially, all models are wrong,
but some are useful”
– George Box

Machine Learning Workflow

Train: Iterate till you find the best model

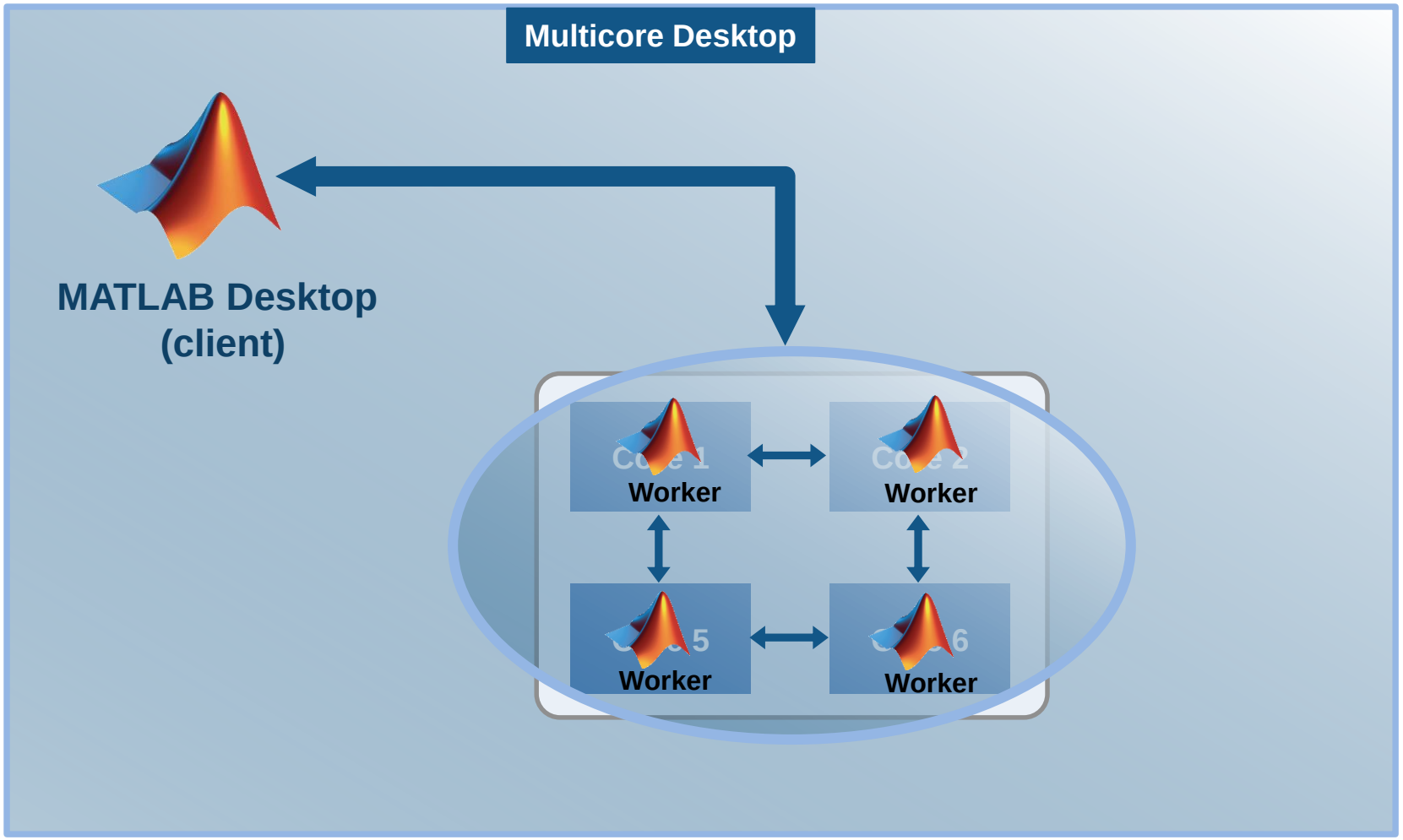


Predict: Integrate trained models into applications



Parallel Computing Paradigm

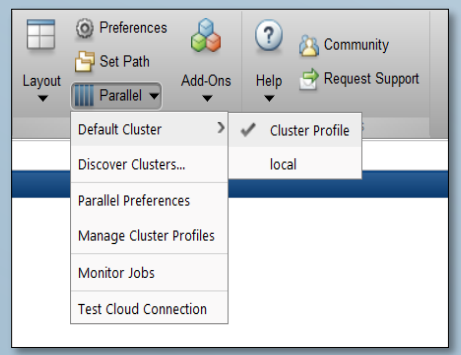
Multicore Desktops



MATLAB multicore

Parallel Computing Paradigm

Cluster Hardware

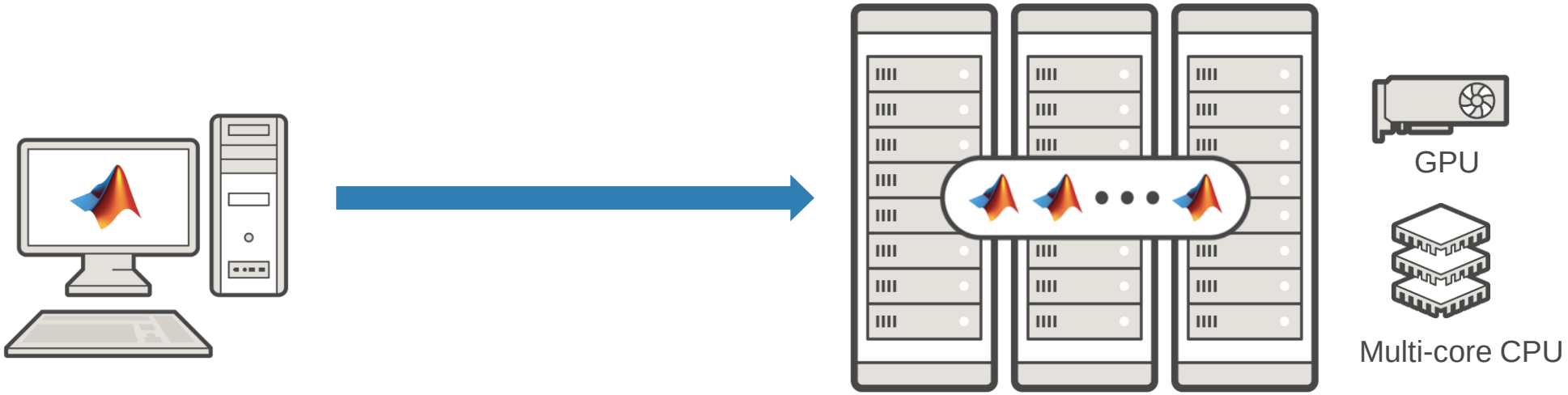


MATLAB Desktop (client)

Cluster of computers

Worker

Migrate execution to a cluster environment



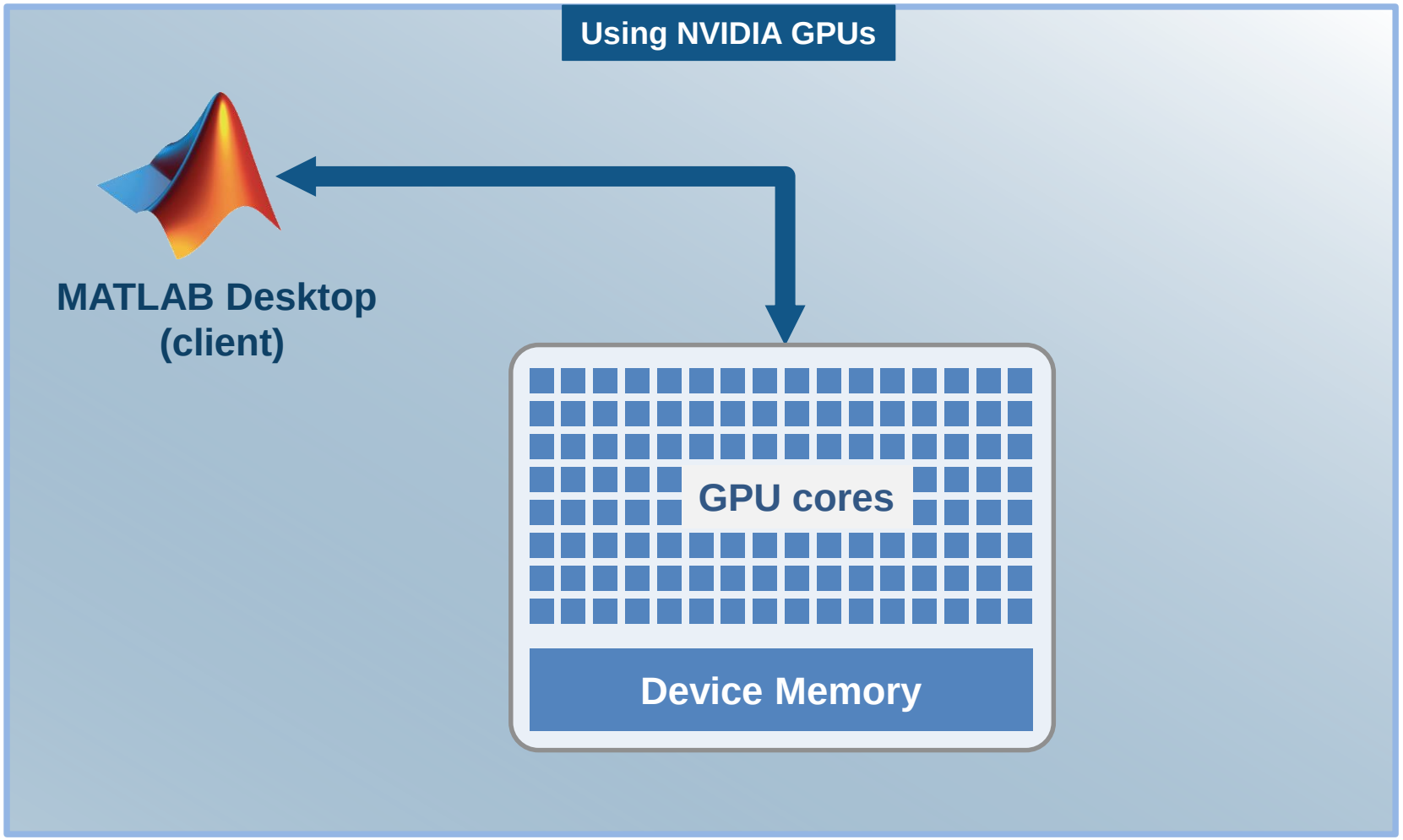
MATLAB

Parallel Computing Toolbox

MATLAB Distributed Computing Server

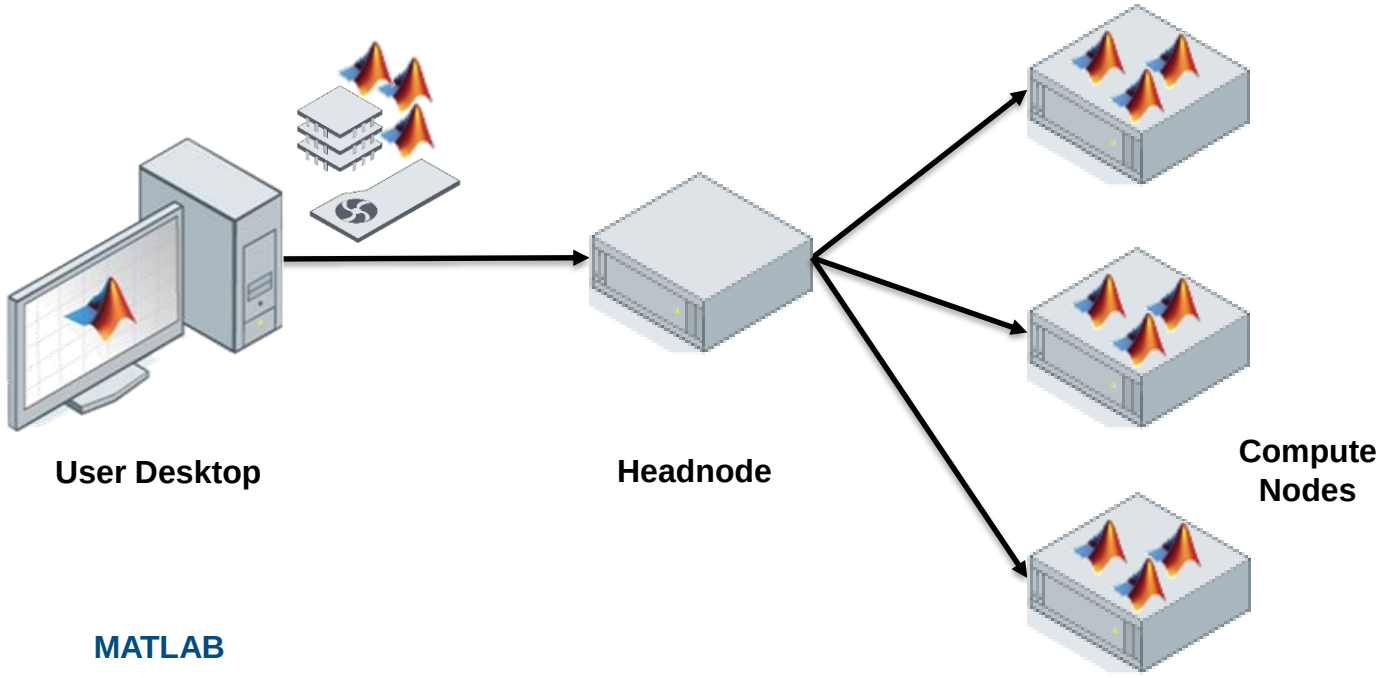
Parallel Computing Paradigm

NVIDIA GPUs



Cluster Computing Paradigm

- Prototype on the desktop
- Integrate with existing infrastructure
- Access directly through MATLAB



Parallel Computing with MATLAB – Beyond PARFOR

Well-known features

- parallel-enabled toolboxes
- `parfor`
- `gpuArray`

Full spectrum of support

- batch submission, jobs and tasks
  `batch, createJob, createTask`
- asynchronous queue for feval
  `parfeval`
- parallel support for big data
  `tall, mapreduce`
- distributed arrays (“global arrays”)
  `distributed, codistributed`
- message passing
  `labSend, labReceive`

tutorials

Parallel-enabled Toolboxes (MATLAB® Product Family)

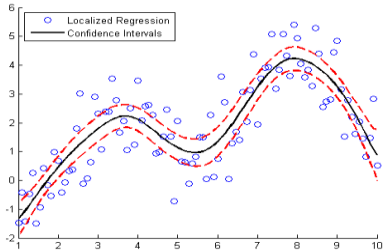
Enable parallel computing support by setting a flag or preference

Image Processing

Batch Image Processor, Block Processing, GPU-enabled functions

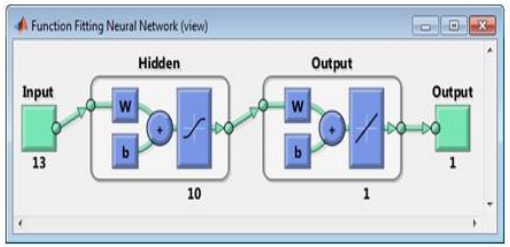
Statistics and Machine Learning

Resampling Methods, k-Means clustering, GPU-enabled functions



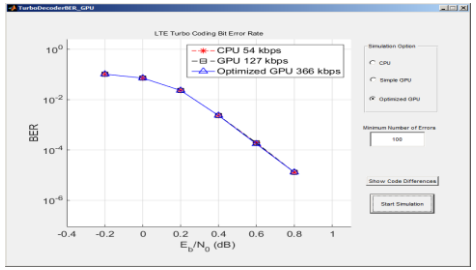
Neural Networks

Deep Learning, Neural Network training and simulation



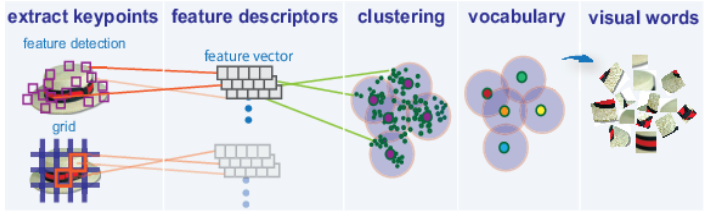
Signal Processing and Communications

GPU-enabled FFT filtering, cross correlation, BER



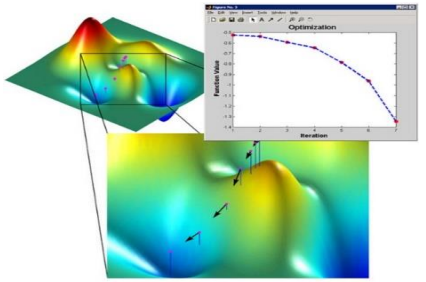
Computer Vision

Parallel-enabled functions in bag-of-words workflow



Optimization

Parallel estimation of gradients

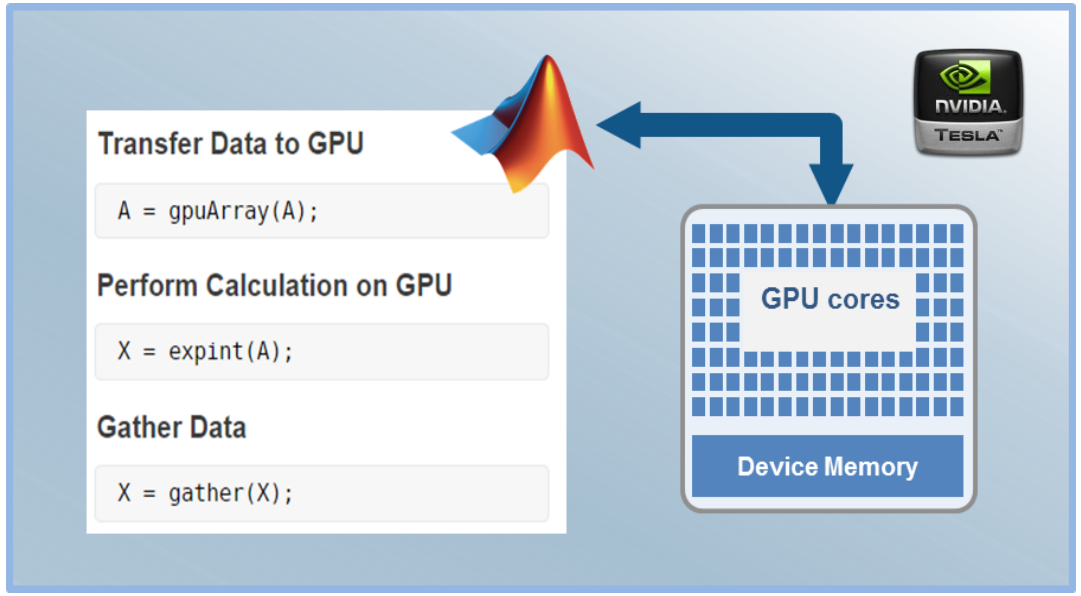


[Other parallel-enabled Toolboxes]

Speed-up MATLAB code with NVIDIA GPUs

- Ideal Problems
  - Massively Parallel and/or Vectorized operations
  - Computationally Intensive
- 300+ GPU-enabled MATLAB functions
  - Enable existing MATLAB code to run on GPUs
  - Support for sparse matrices on GPUs
- Additional GPU-enabled Toolboxes
  - Neural Networks
  - Image Processing
  - Signal Processing

..... Learn More

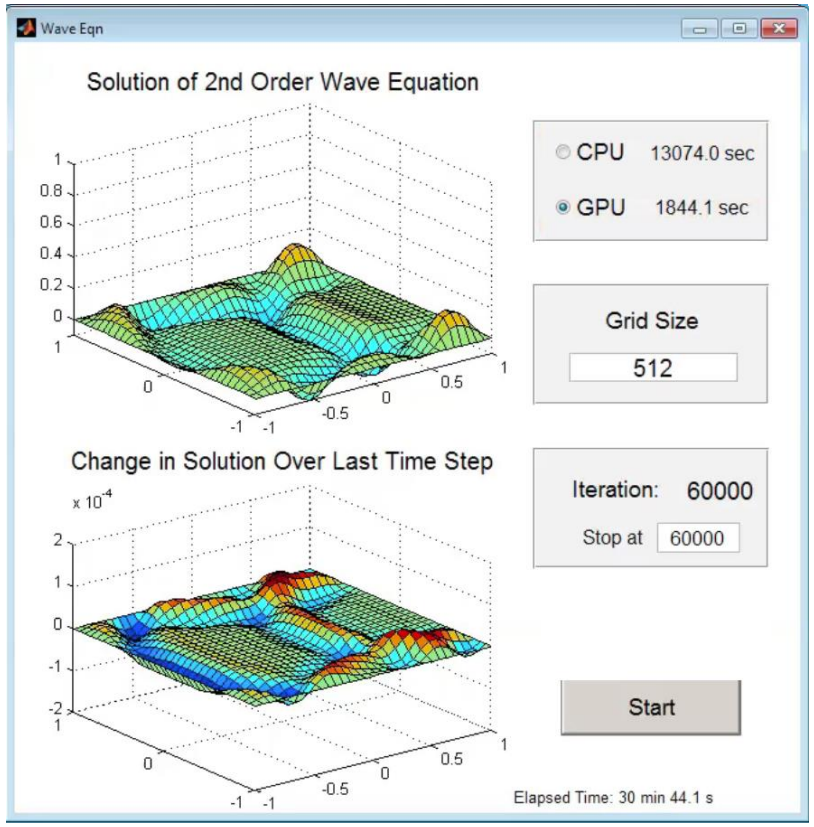


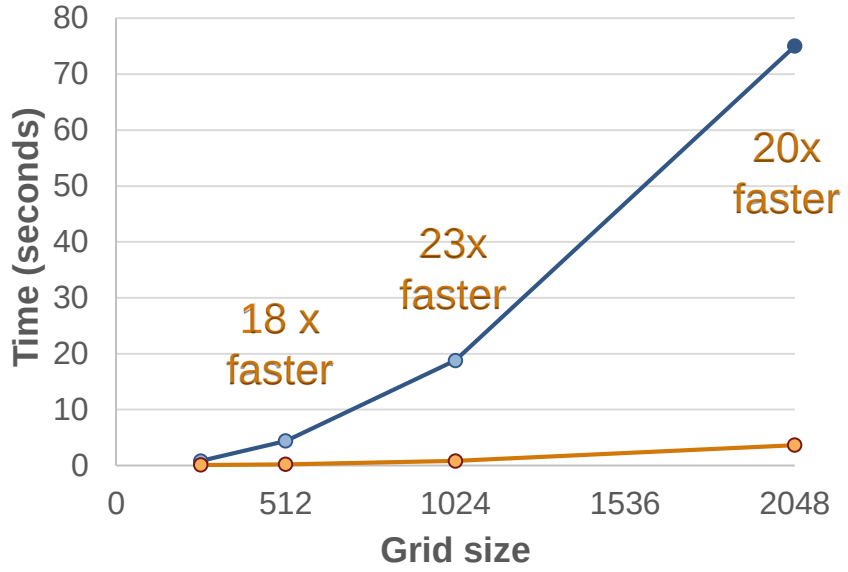
Create a Sparse gpuArray

```
i = [900 1000];
j = [900 1000];
v = [10 100];
s = sparse(i,j,v,1500,1500);
g = gpuArray(s);
```

Run Same Code on CPU and GPU

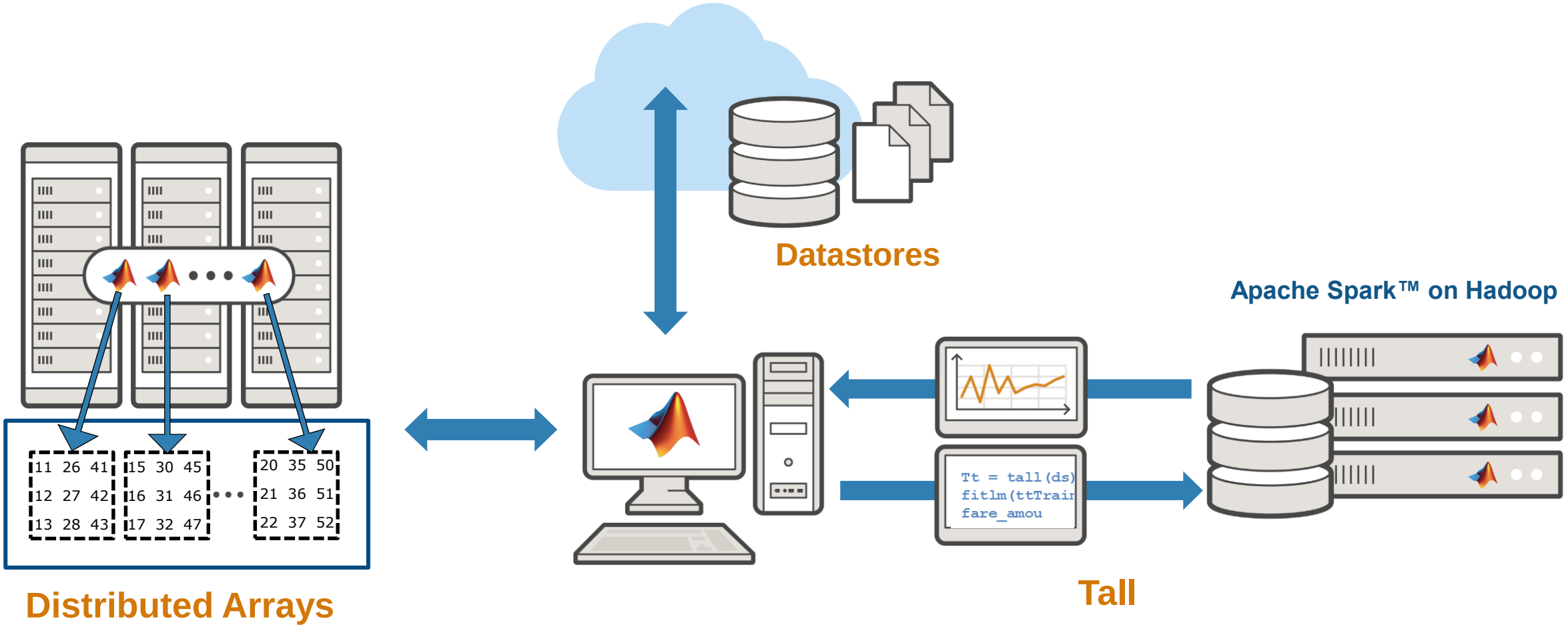
Solving 2D Wave Equation





CPU	GPU
Intel(R) Xeon(R) W3550 3.06GHz 4 cores memory bandwidth 25.6 Gb/s	NVIDIA Tesla K20c 706MHz 2496 cores memory bandwith 208 Gb/s

Big Data capabilities in MATLAB



Big Data capabilities in MATLAB

ACCESS

Access more data and collections of files than fit in memory

Large collections of data files

- datastore
- support for HDFS

PROCESS AND ANALYZE

Adapt traditional processing tools or learn new tools to work with Big Data

Statistical analysis

- tall arrays
- distributed arrays and overloaded functions

Signal processing

- distributed arrays and overloaded functions

Deep Learning

- GPU

SCALE

Scale to compute clusters and Hadoop/Spark

Analysis of large tabular data

- tall arrays

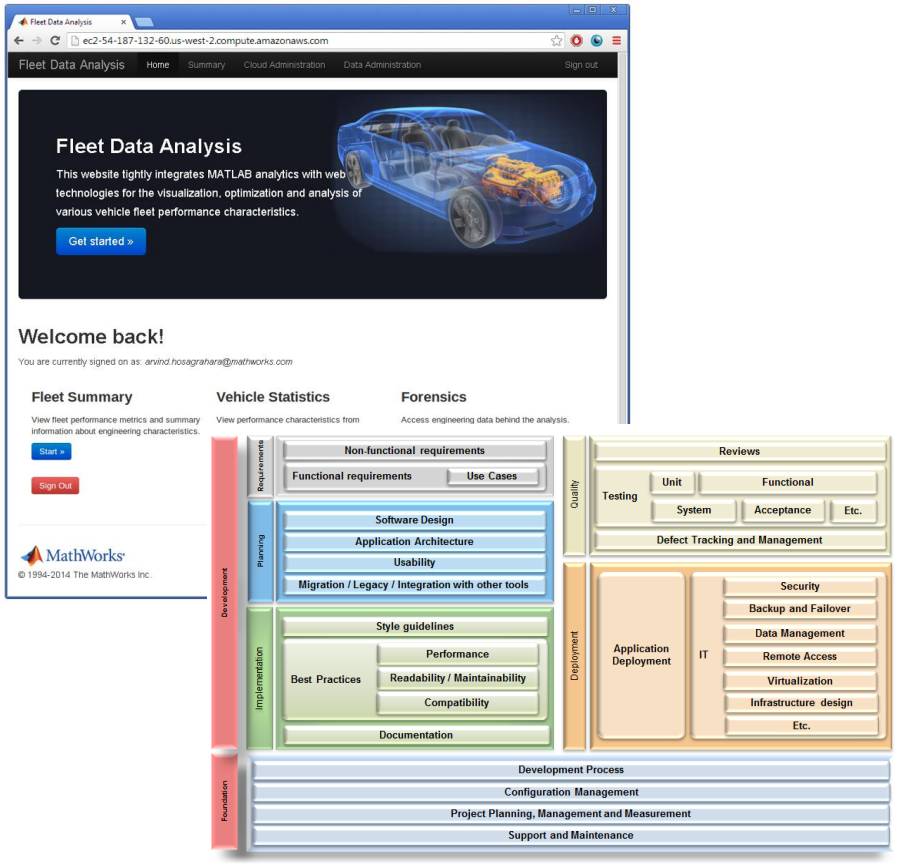
Large simulations of environmental data

- distributed arrays and overloaded functions

Advanced techniques for power users

- MATLAB API for Spark
- mapreduce
- labSend / labReceive

MathWorks Services



- Consulting
  - Integration
  - Data analysis/visualization
  - Unify workflows, models, data

www.mathworks.com/services/consulting/

- Training
  - Classroom, online, on-site
  - Data Processing, Visualization, Deployment, Parallel Computing

www.mathworks.com/services/training/