

MATLAB EXPO 2017

KOREA

4월 27일, 서울

등록 하기 matlabexpo.co.kr

엔터프라이즈 시스템에서의 빅데이터 애널리틱스 애플리케이션 구축을 위한 MATLAB 기능

성호현 차장

Senior Application Engineer

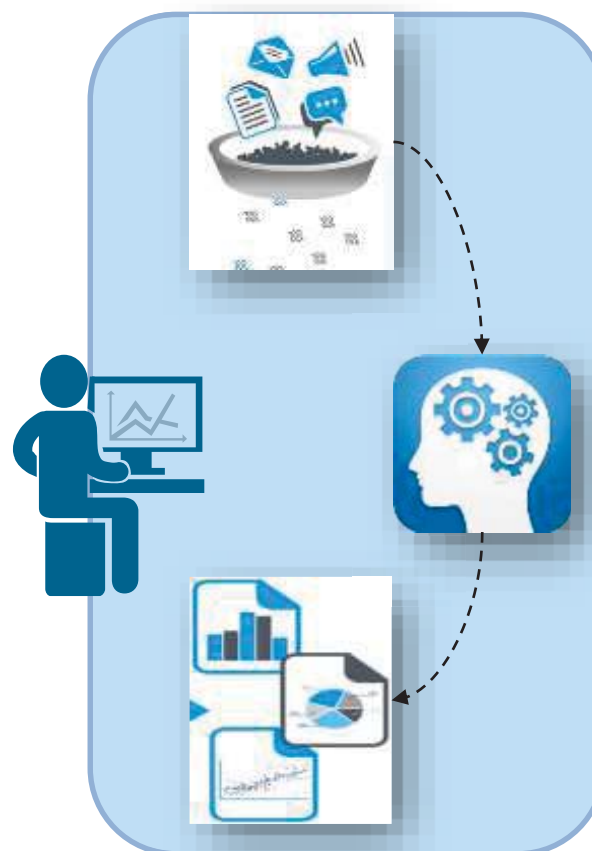
The MathWorks Korea

Data Analytics Workflow



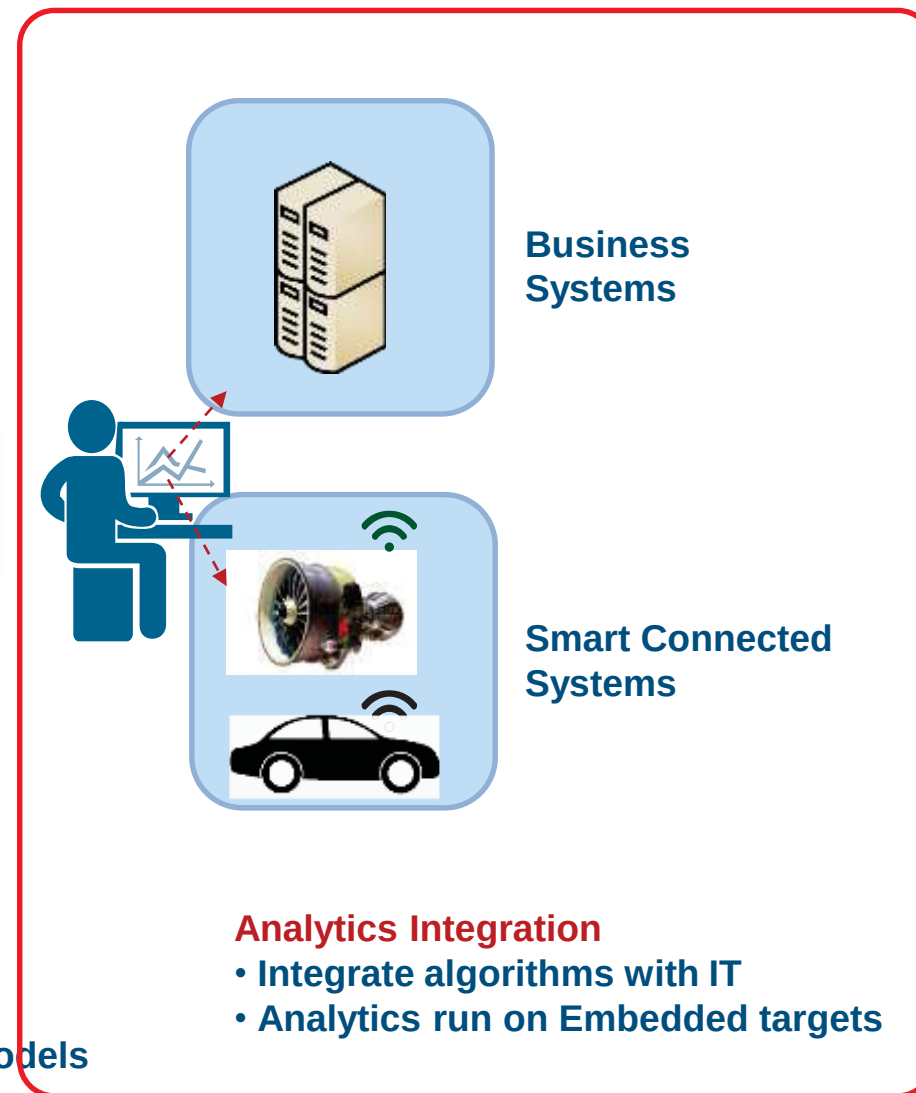
Data Acquisition

- Engineering, Scientific, and Field
- Business and Transactional



Data Analytics

- Data Pre-processing
- Feature Extraction
- Building algorithms, math models
- Making business decisions



Analytics Integration

- Integrate algorithms with IT
- Analytics run on Embedded targets

Key Takeaways

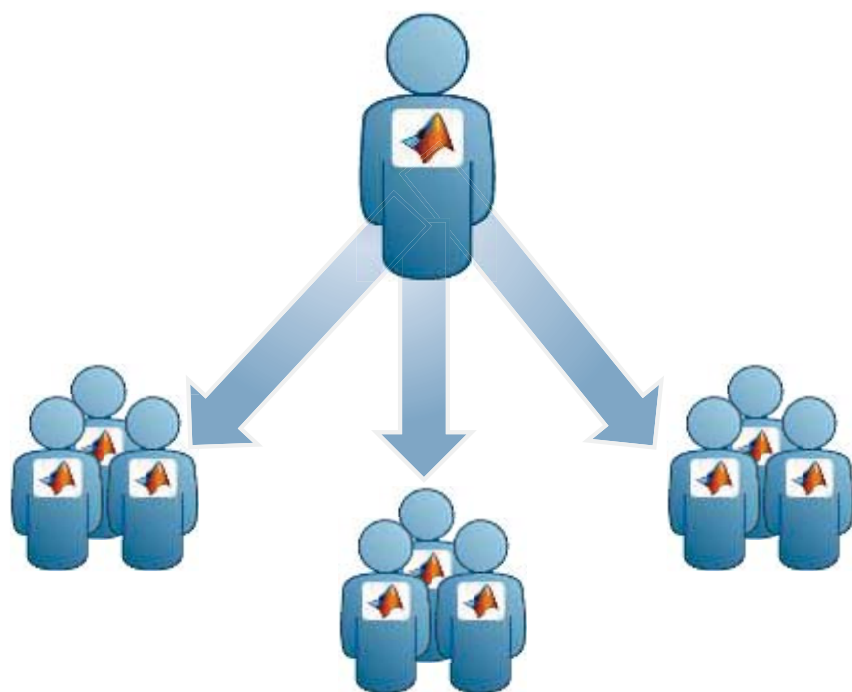
1. Distribute applications to non-MATLAB users royalty-free.
2. Integrate MATLAB functions into existing workflows and development platforms.
3. Deploy MATLAB Analytics for Big Data on Hadoop enabled Spark Clusters .
4. Deploy MATLAB applications to service simultaneous user requests enterprise-wide via web or cloud frameworks.

Challenges

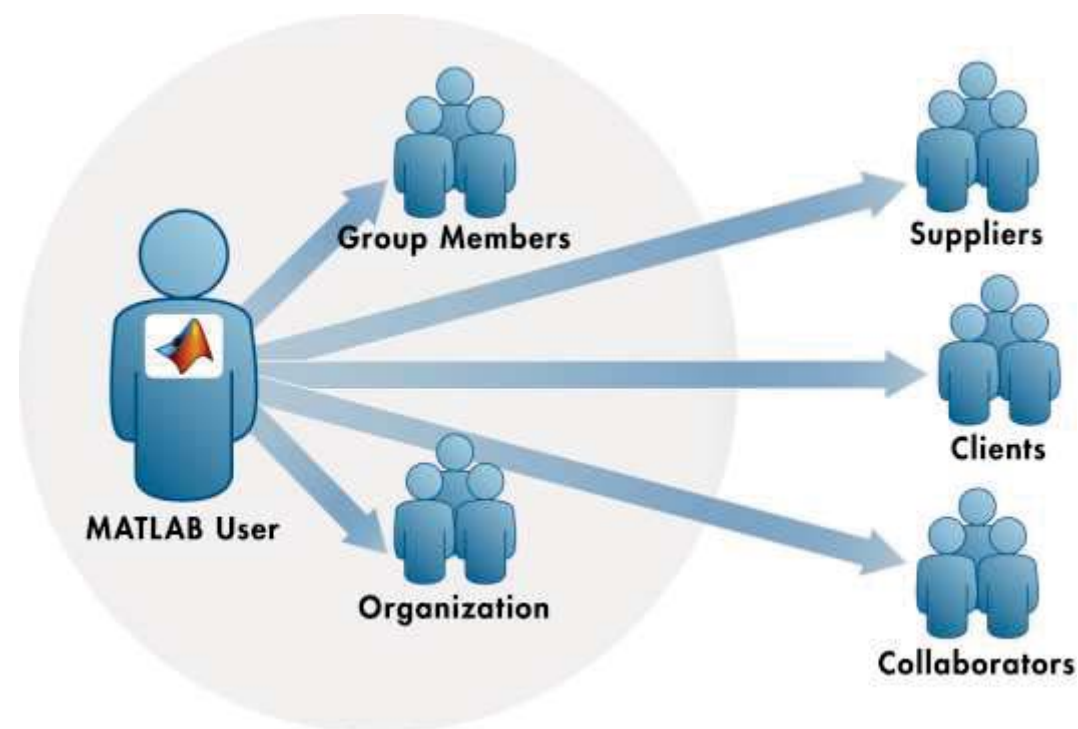
- Multiple internal and external consumers of MATLAB algorithms
- Challenging and time consuming to re-code MATLAB algorithms for integration into IT frameworks
 - Development resources are scarce and time-to-market is short
- Company priority to deploy solutions to enterprise scale web or cloud frameworks
 - Scale application to serve large numbers of simultaneous requests

MATLAB Programs Can be Shared With Anyone

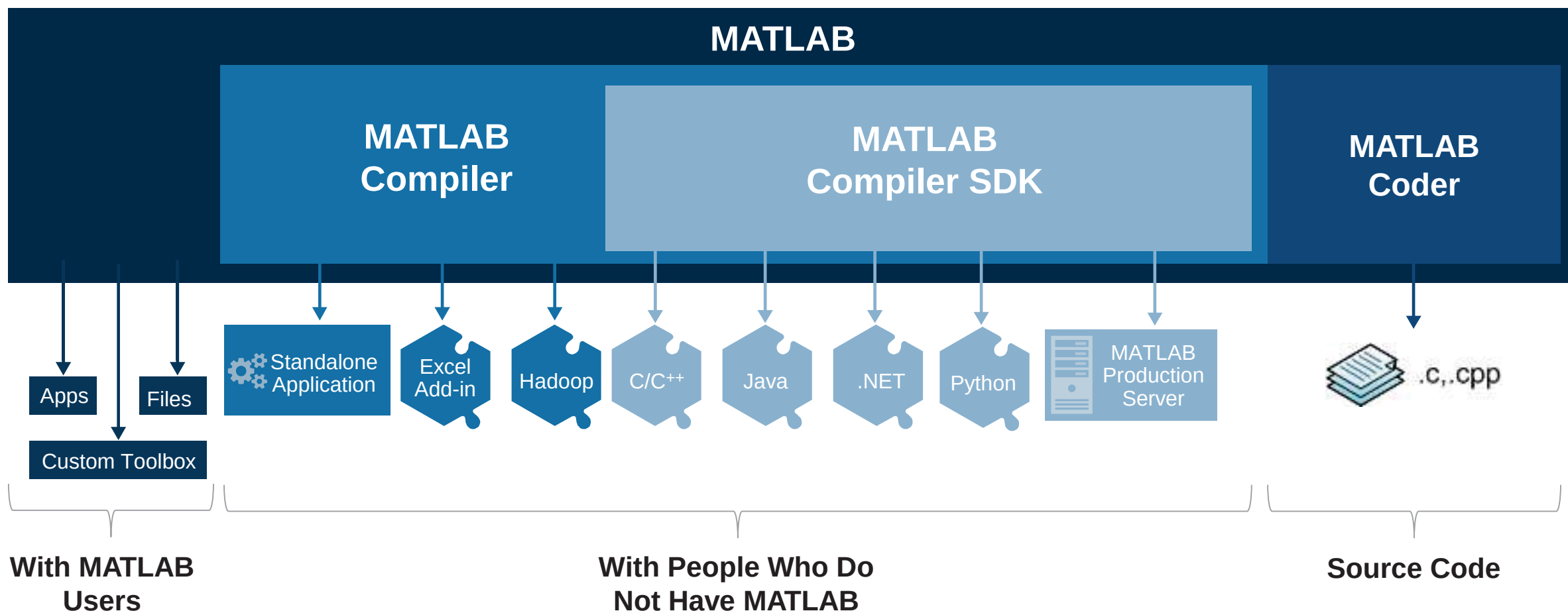
Share With Other MATLAB Users



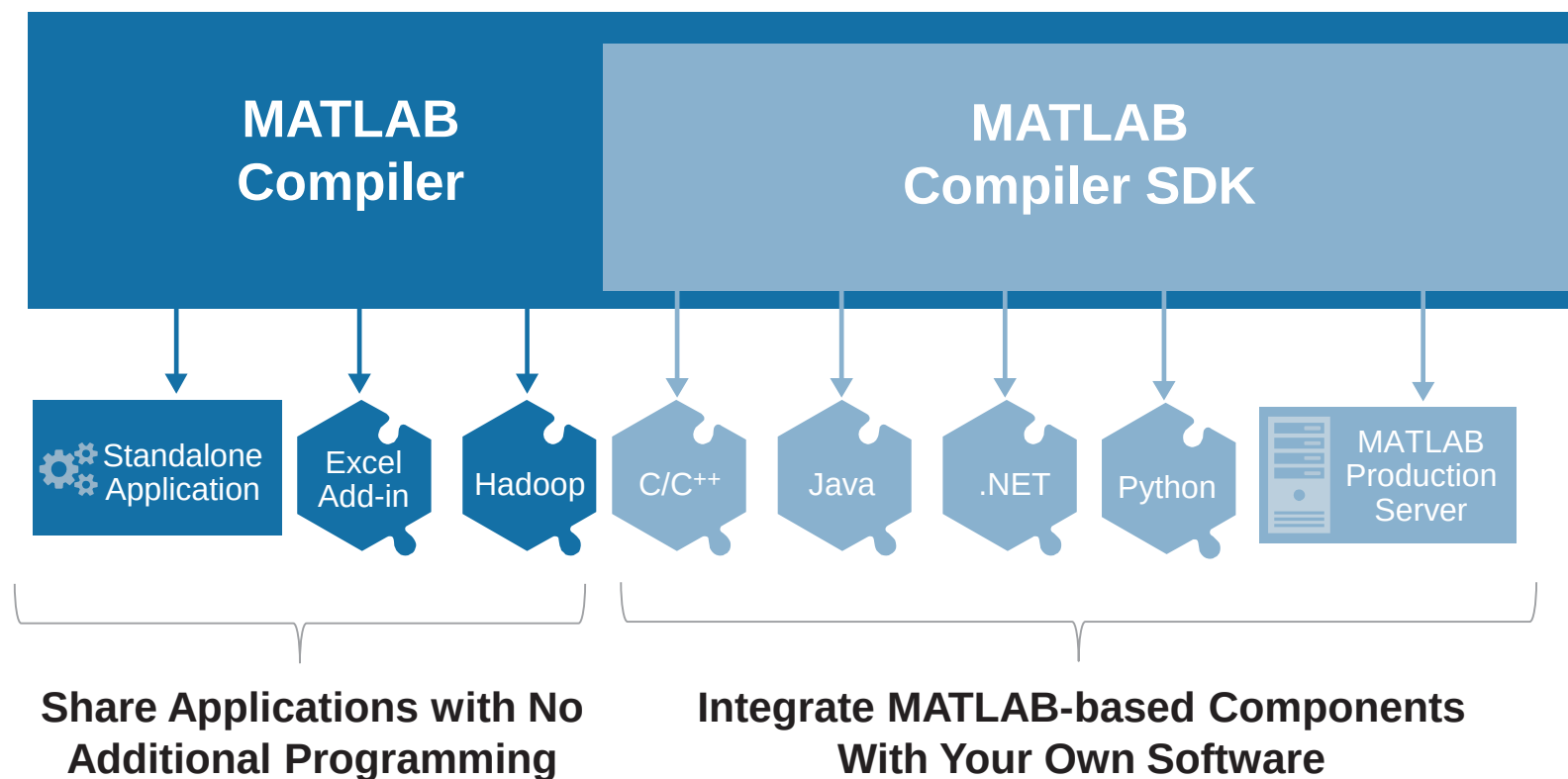
Share With People Who do Not Have MATLAB



Write Your Programs Once Then Share To Different Targets

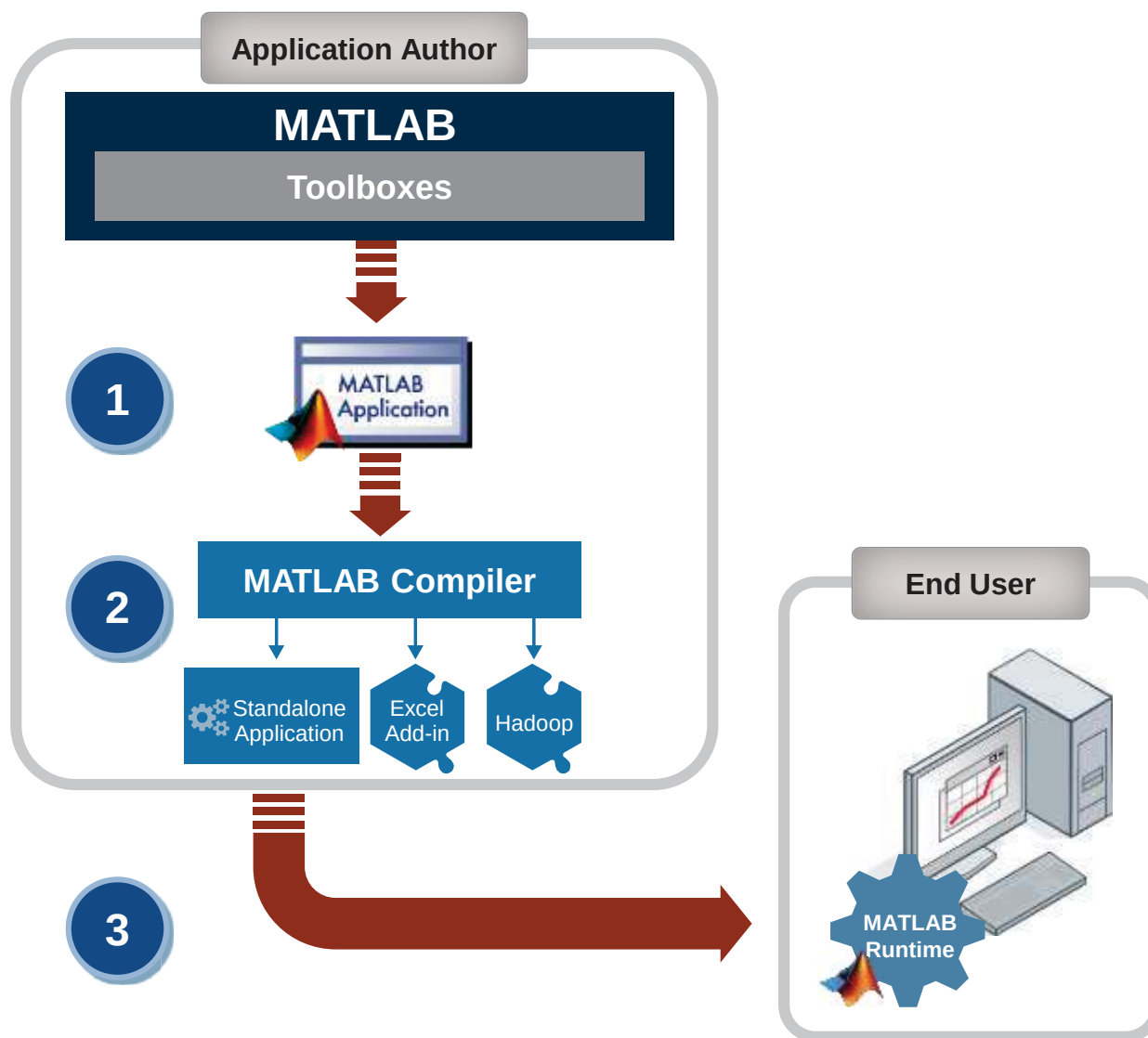


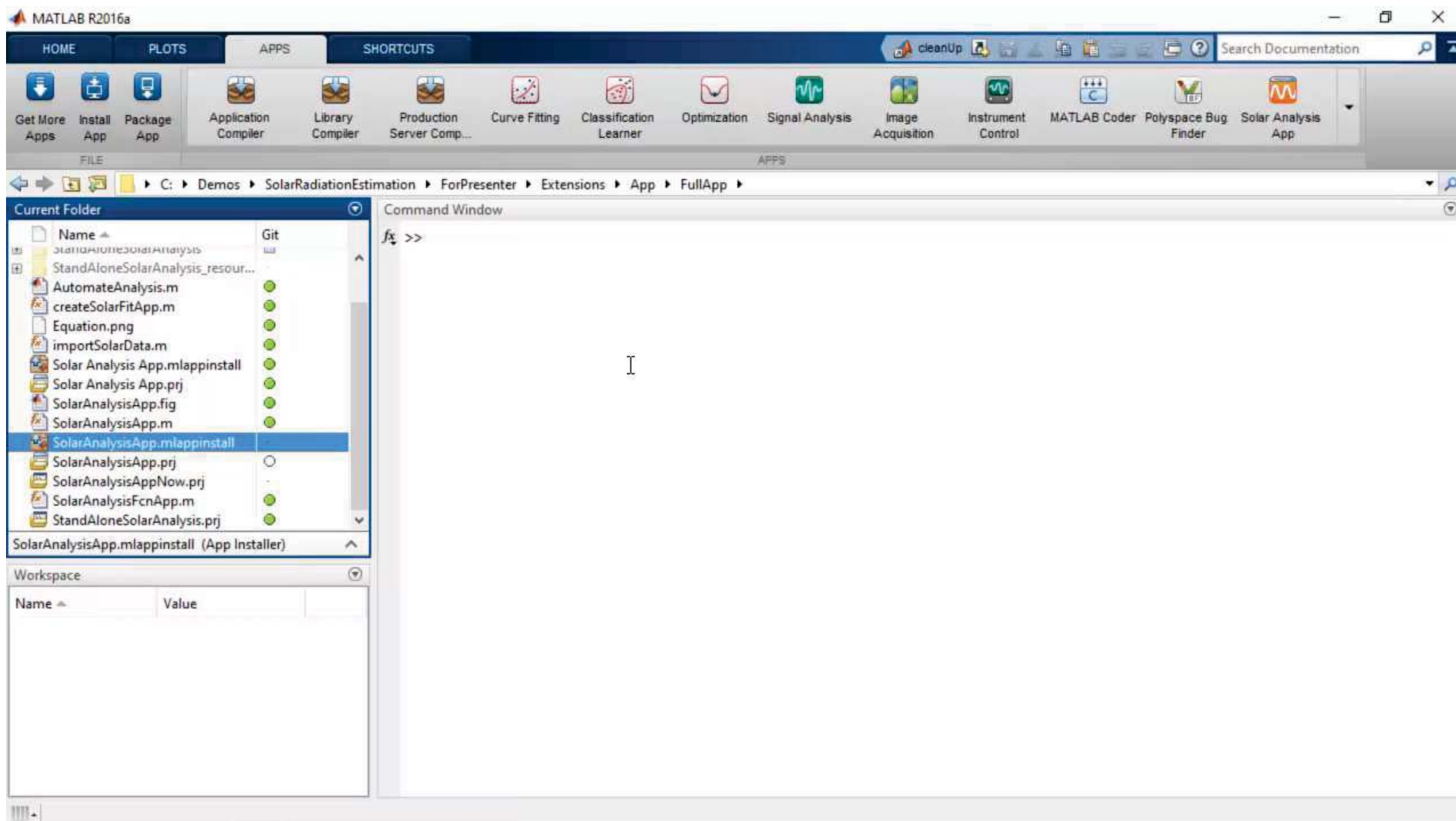
Share with People Who Do Not Have MATLAB



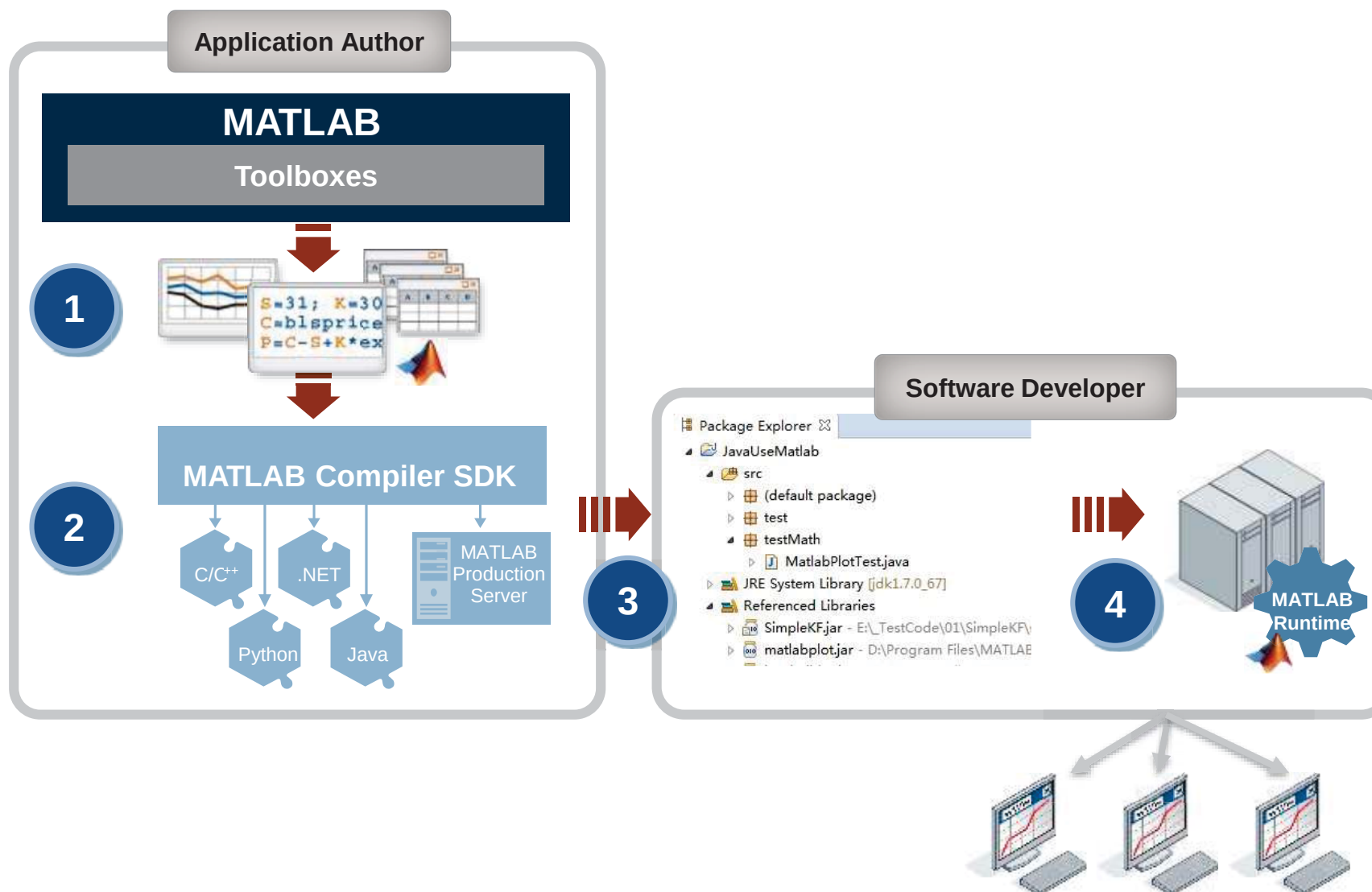
- Royalty-free Sharing
- IP Protection via Encryption

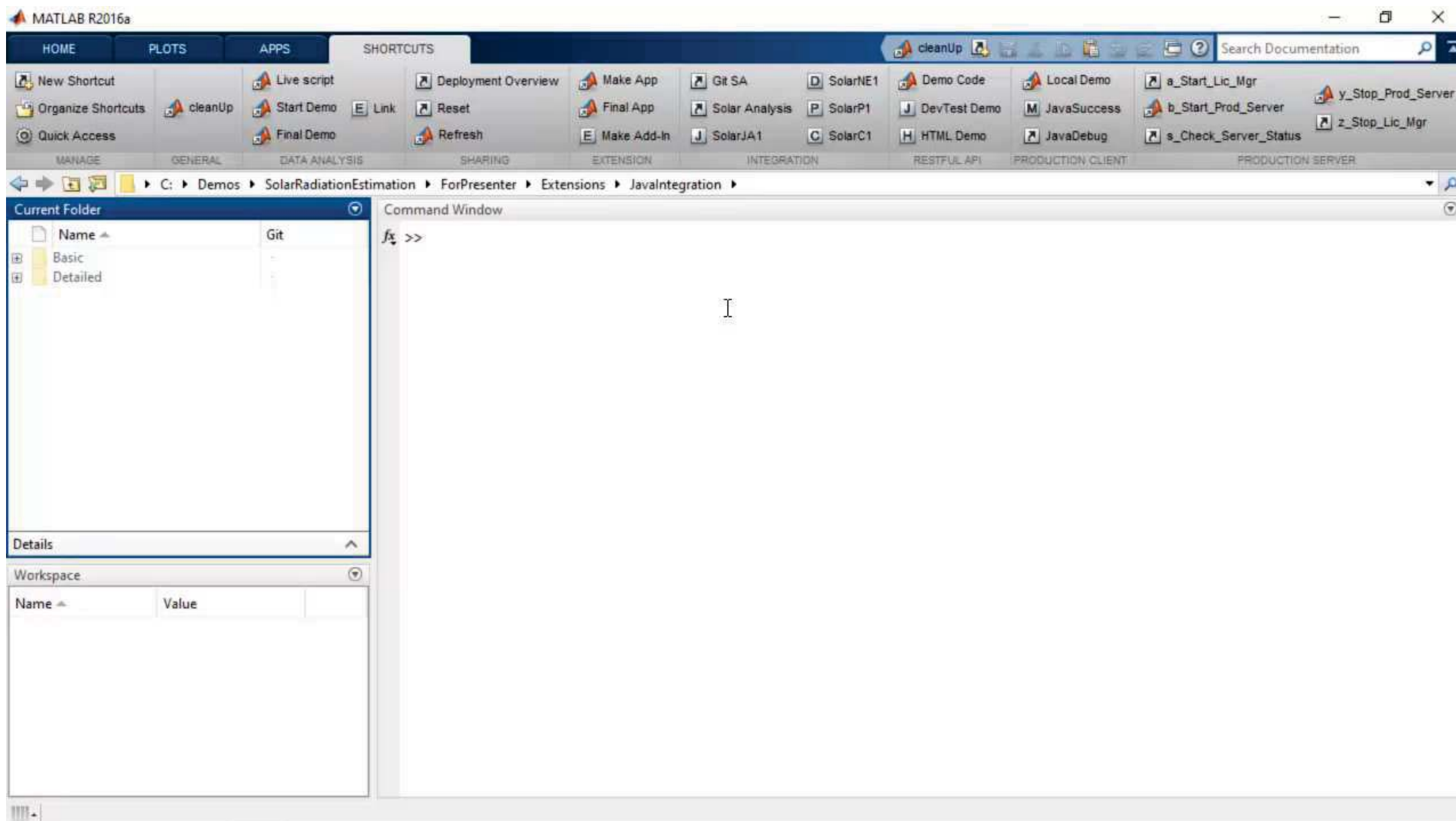
Share Applications Built Completely in MATLAB



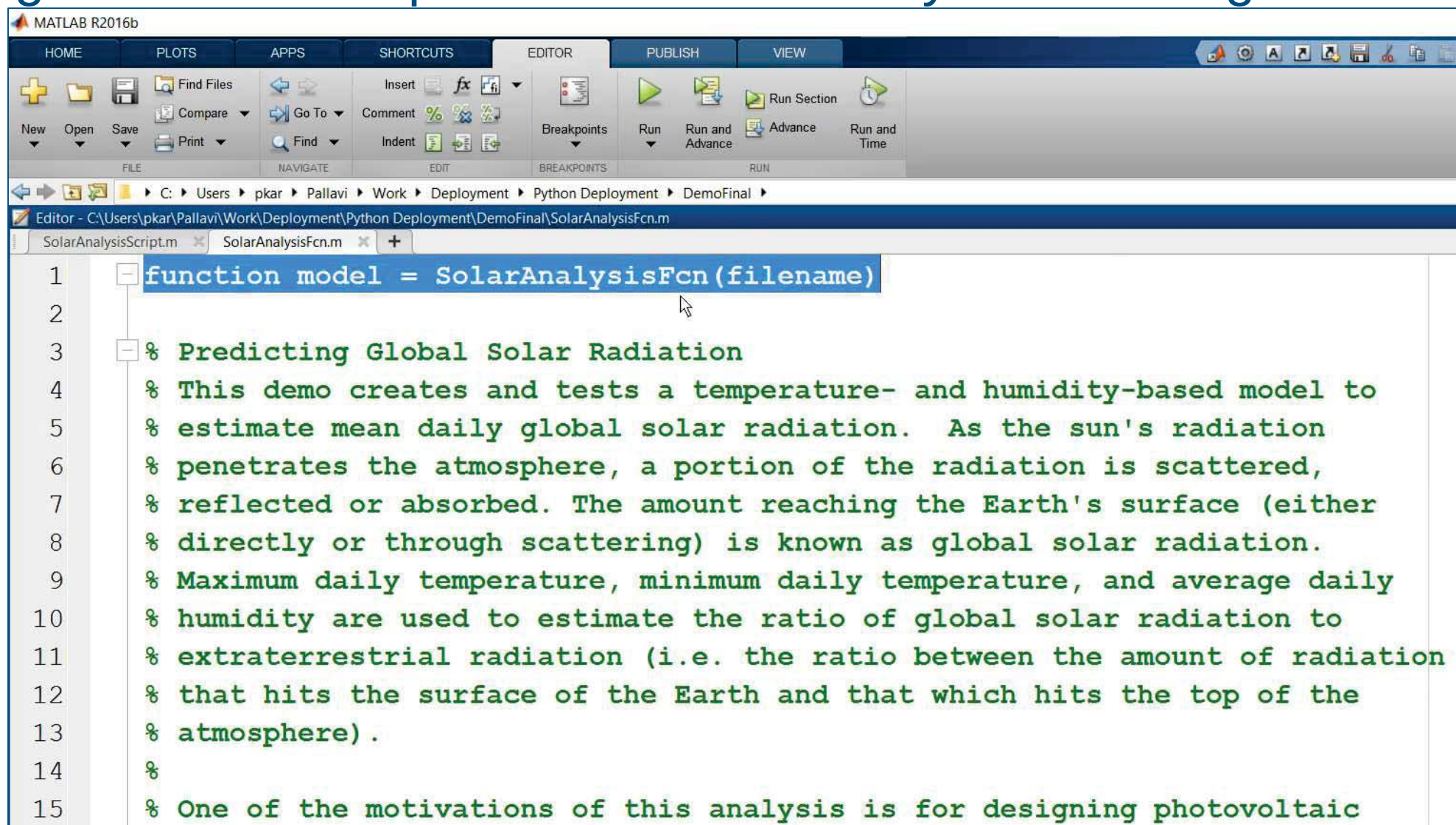


Integrate MATLAB-based Components With Your Own Software





Using MATLAB Compiler SDK to create Python Packages



The image shows the MATLAB R2016b interface. The top toolbar includes tabs for HOME, PLOTS, APPS, SHORTCUTS, EDITOR, PUBLISH, and VIEW. The EDITOR tab is active, showing a file explorer on the left and a code editor on the right. The code editor displays a function definition for `SolarAnalysisFcn`. The function signature is `function model = SolarAnalysisFcn(filename)`. The function body contains several comments explaining the purpose of the function: predicting global solar radiation based on temperature and humidity data. The code is as follows:

```

1 function model = SolarAnalysisFcn(filename)
2
3 % Predicting Global Solar Radiation
4 % This demo creates and tests a temperature- and humidity-based model to
5 % estimate mean daily global solar radiation. As the sun's radiation
6 % penetrates the atmosphere, a portion of the radiation is scattered,
7 % reflected or absorbed. The amount reaching the Earth's surface (either
8 % directly or through scattering) is known as global solar radiation.
9 % Maximum daily temperature, minimum daily temperature, and average daily
10 % humidity are used to estimate the ratio of global solar radiation to
11 % extraterrestrial radiation (i.e. the ratio between the amount of radiation
12 % that hits the surface of the Earth and that which hits the top of the
13 % atmosphere).
14 %
15 % One of the motivations of this analysis is for designing photovoltaic

```

MATLAB *and* MATLAB Production Server

is the **easiest** and most **productive** environment to *take your enterprise analytics or IoT solution* from **idea** to **production**



Idea



Production

Why MATLAB Production Server Matters to You



Domain Expert

- ✓ MATLAB Production Server allow you to continue to work in the environment that you love
- ✓ No need to learn another programming language
- ✓ MATLAB Production Server integrates with enterprise IT infrastructure

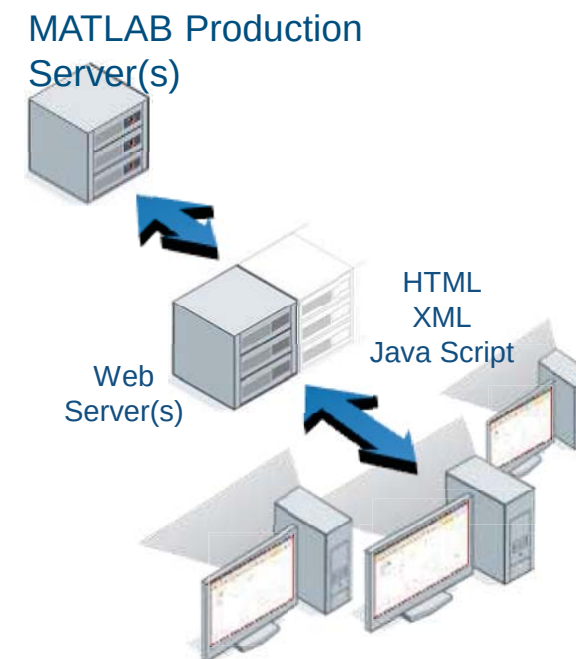


Solution Architect

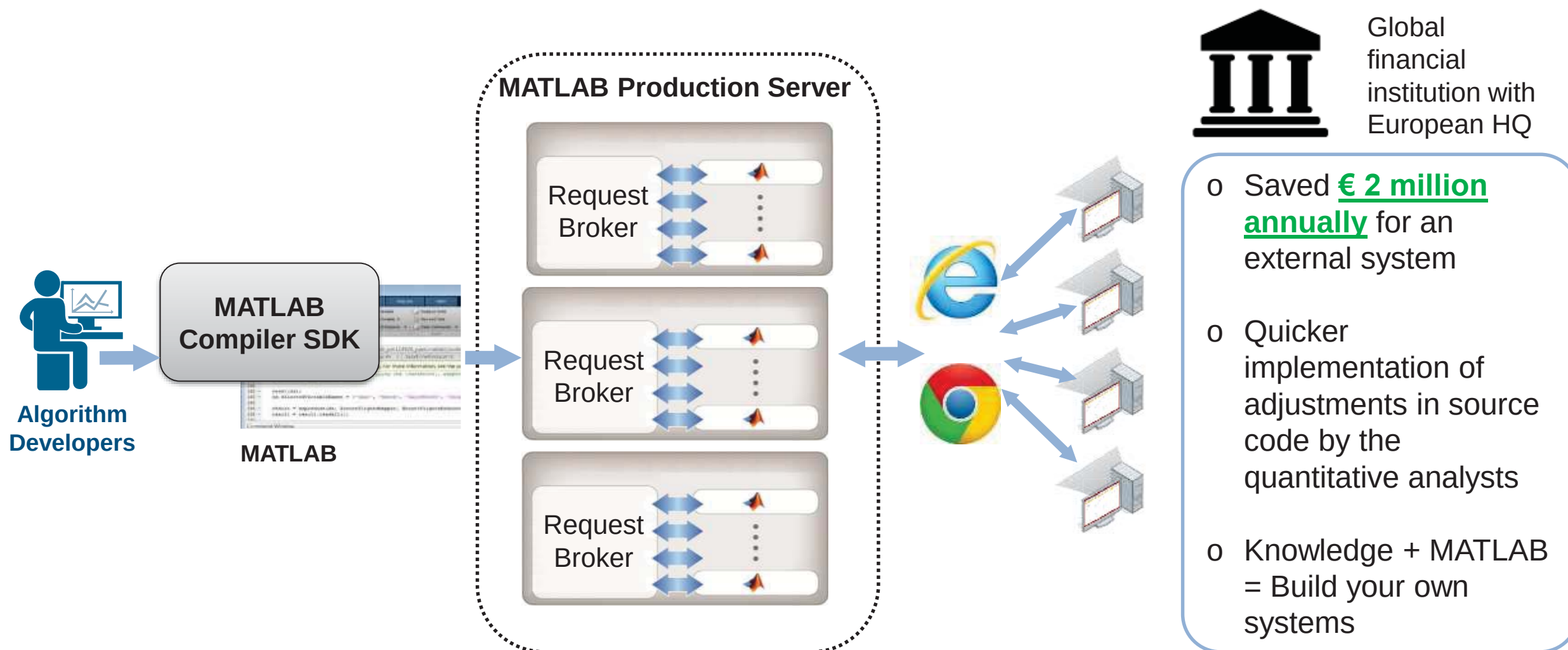
- ✓ MATLAB Production Server integrates MATLAB code into the enterprise IT fabric that you are comfortable with
- ✓ No need to re-code into another programming language
- ✓ Web and cloud friendly architecture

Scale Up with MATLAB Production Server™

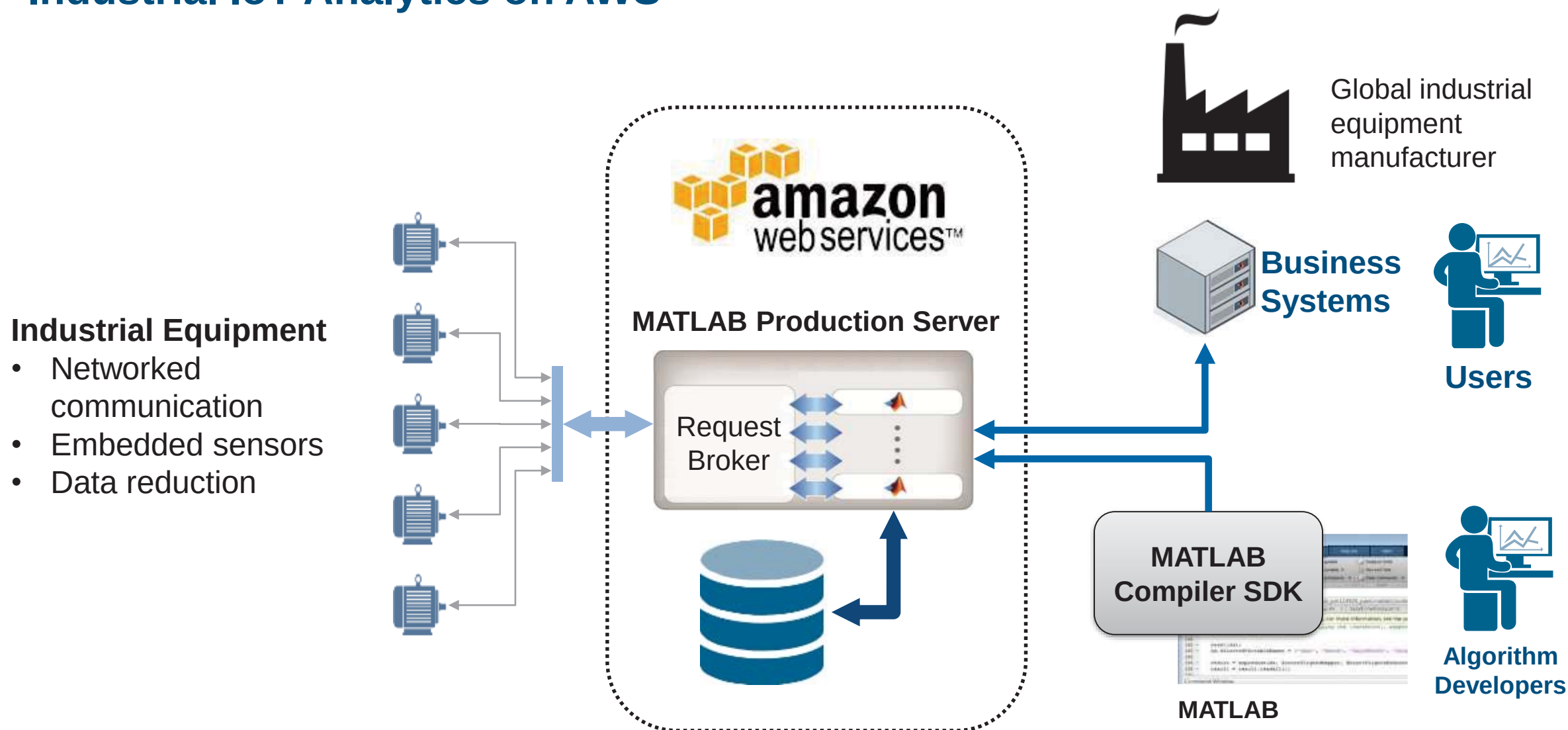
- Directly deploy MATLAB programs into production
 - Centrally manage multiple MATLAB programs and runtime versions
 - Automatically deploy updates without server restarts
 - Most efficient path for creating enterprise applications
- Scalable and reliable
 - Service large numbers of concurrent requests
 - Add capacity or redundancy with additional servers
- Use with web, database and application servers
 - Lightweight client library isolates MATLAB processing
 - Access MATLAB programs using native data types



Customer examples: Financial customer advisory service



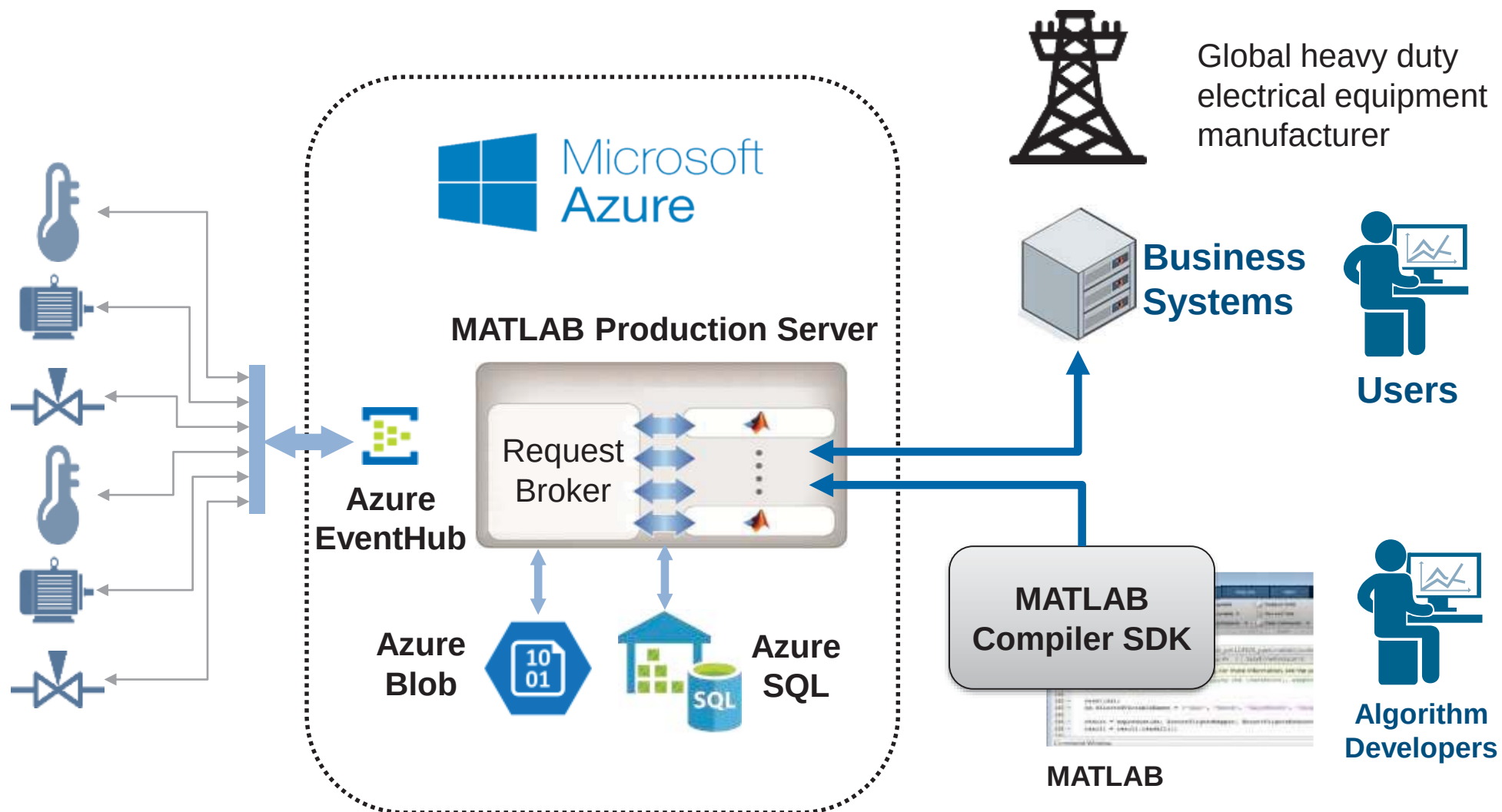
Industrial IoT Analytics on AWS



Building Automation IoT Analytics on Azure

Building/HVAC automation control system

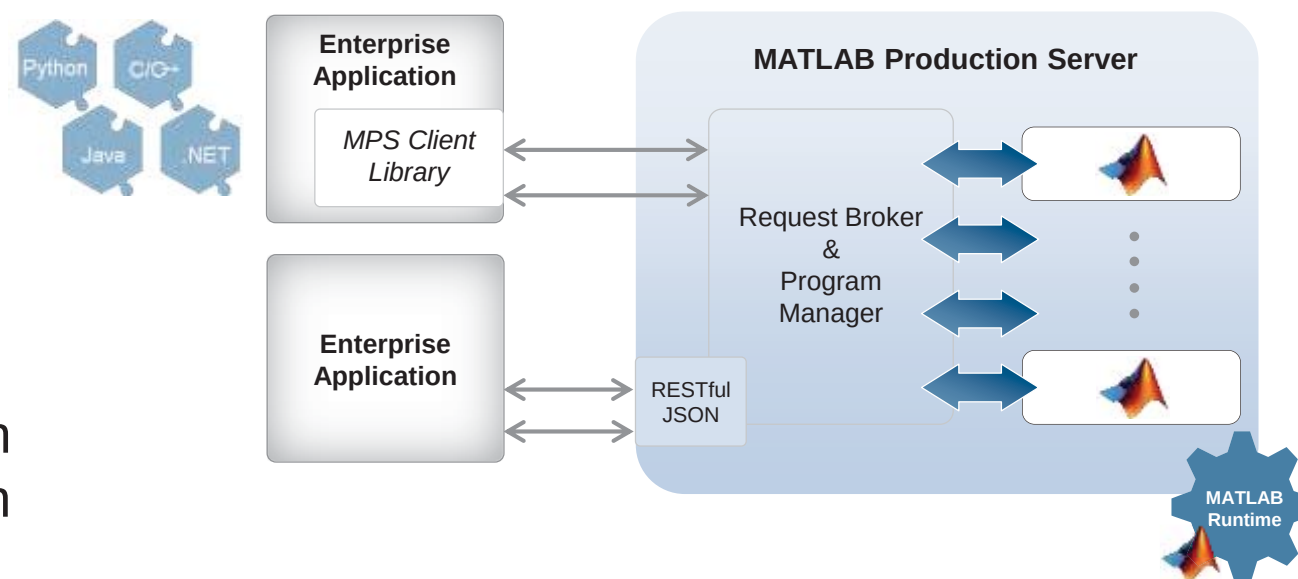
- Variety of sensors and controls
- Networked communication
- Data reduction



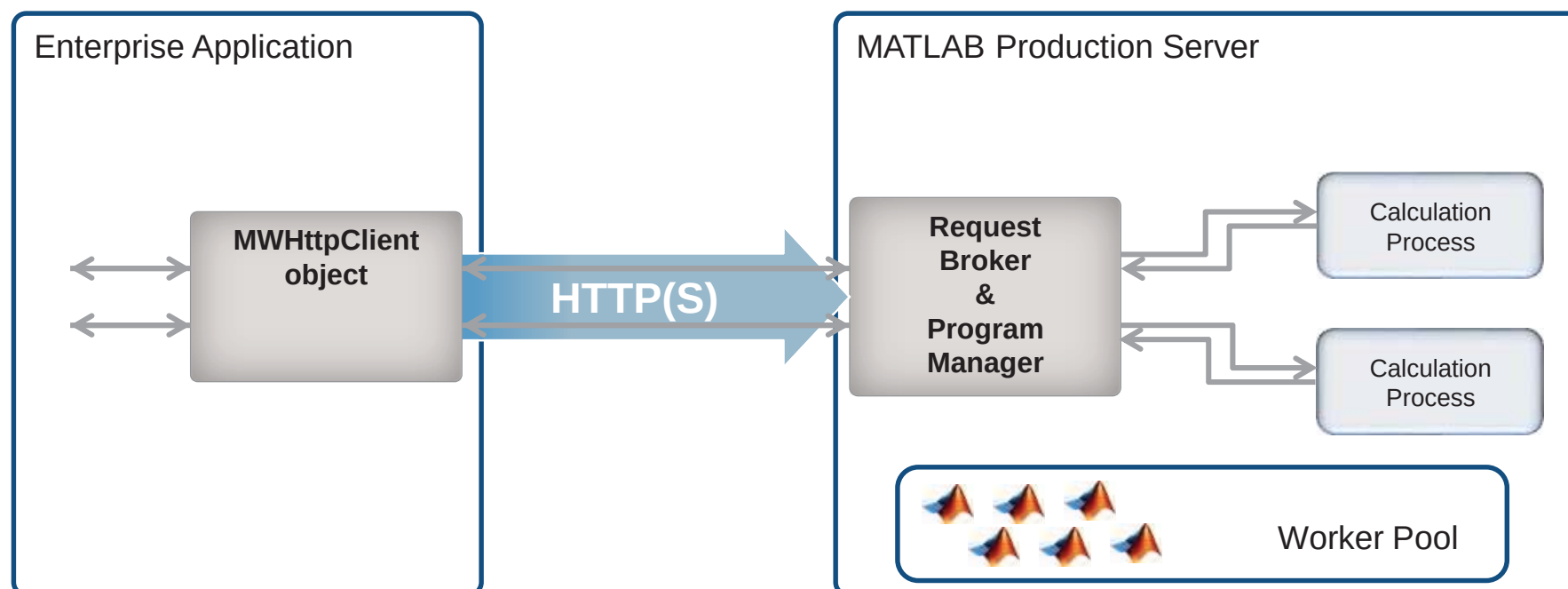
MATLAB Production Server

Enterprise Class Framework For Running Packaged MATLAB Programs

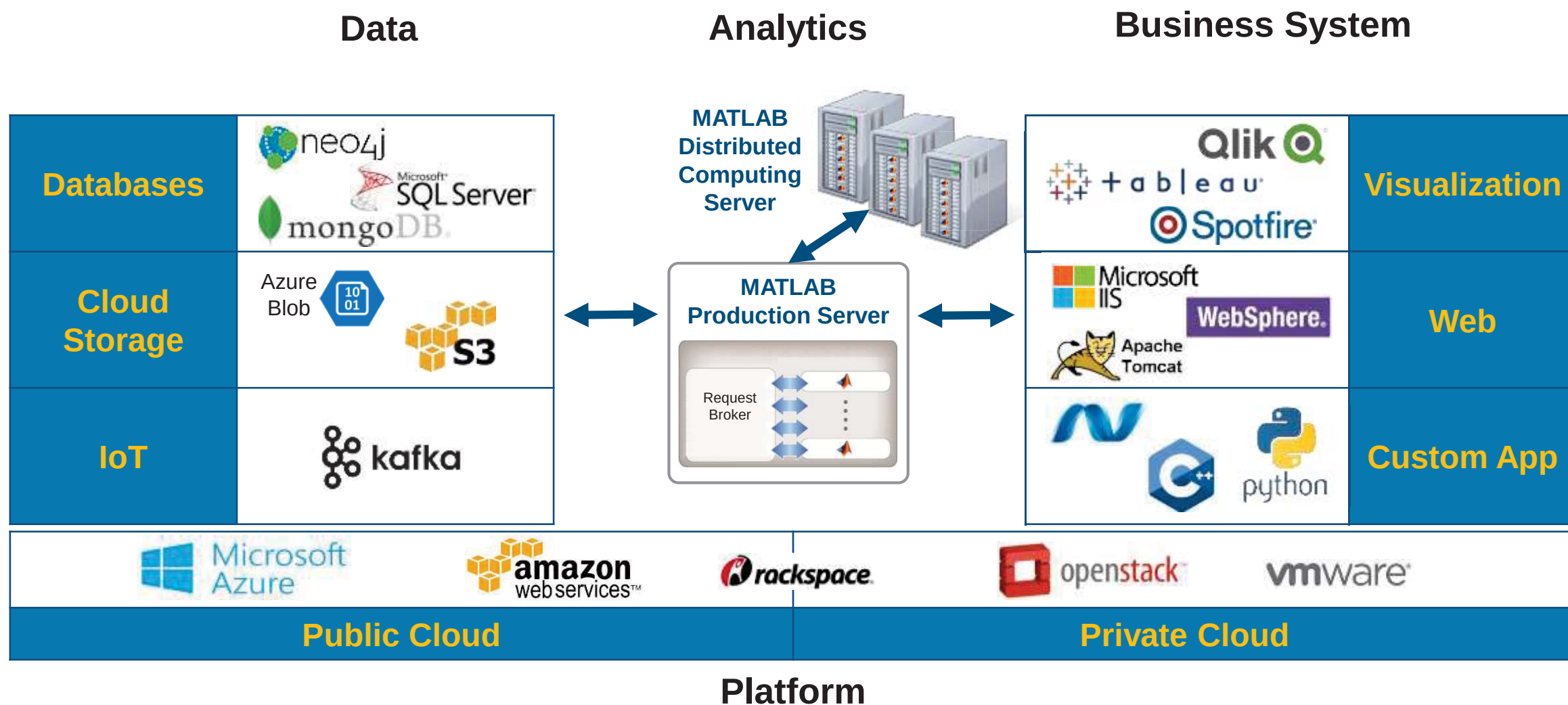
- Server software
 - Manages packaged MATLAB programs and worker pool
- MATLAB Runtime libraries
 - Single server can use runtimes from different releases
- RESTful JSON interface and lightweight client library (C/C++, .NET, Python, and Java)



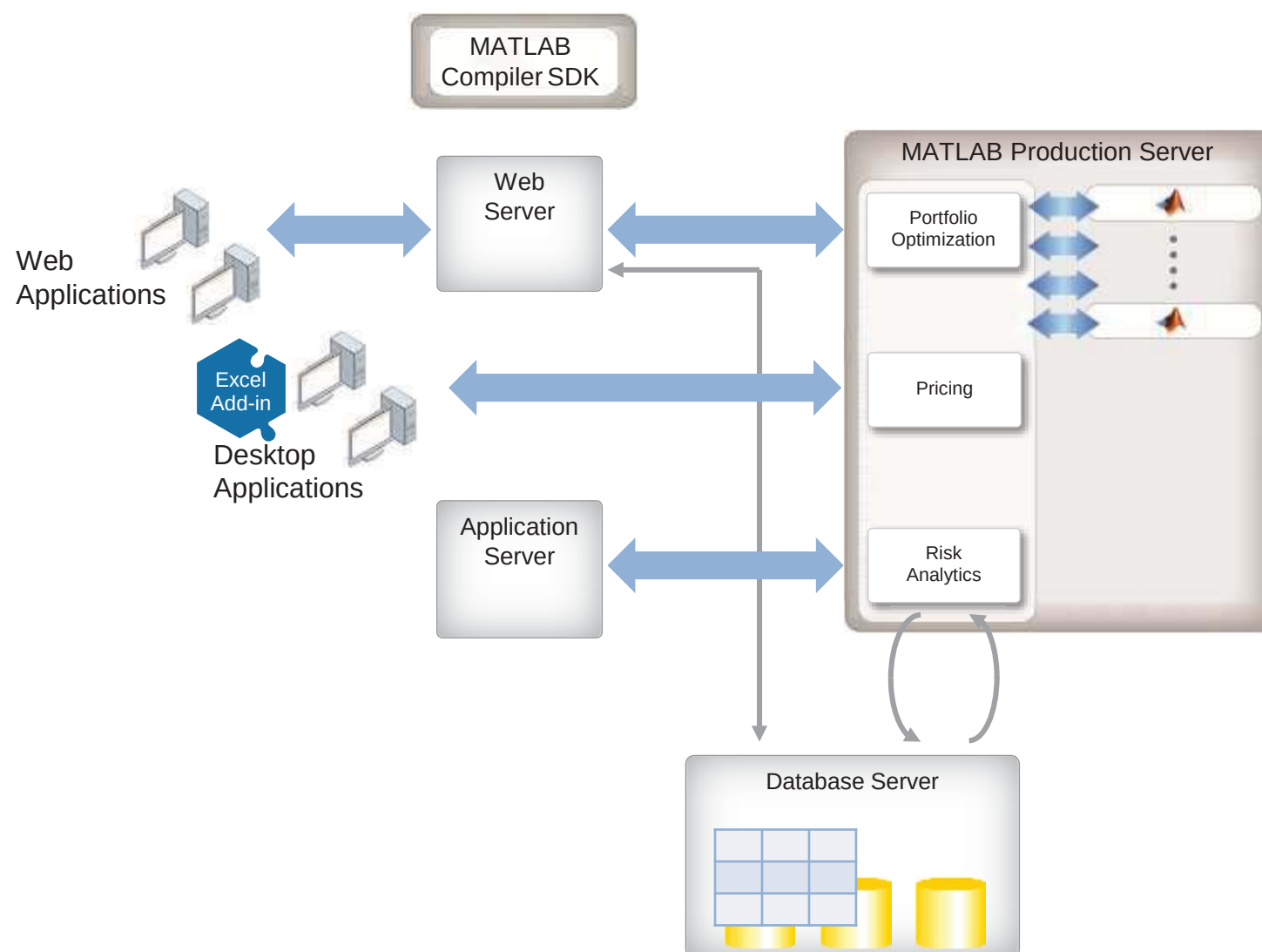
Calling Functions



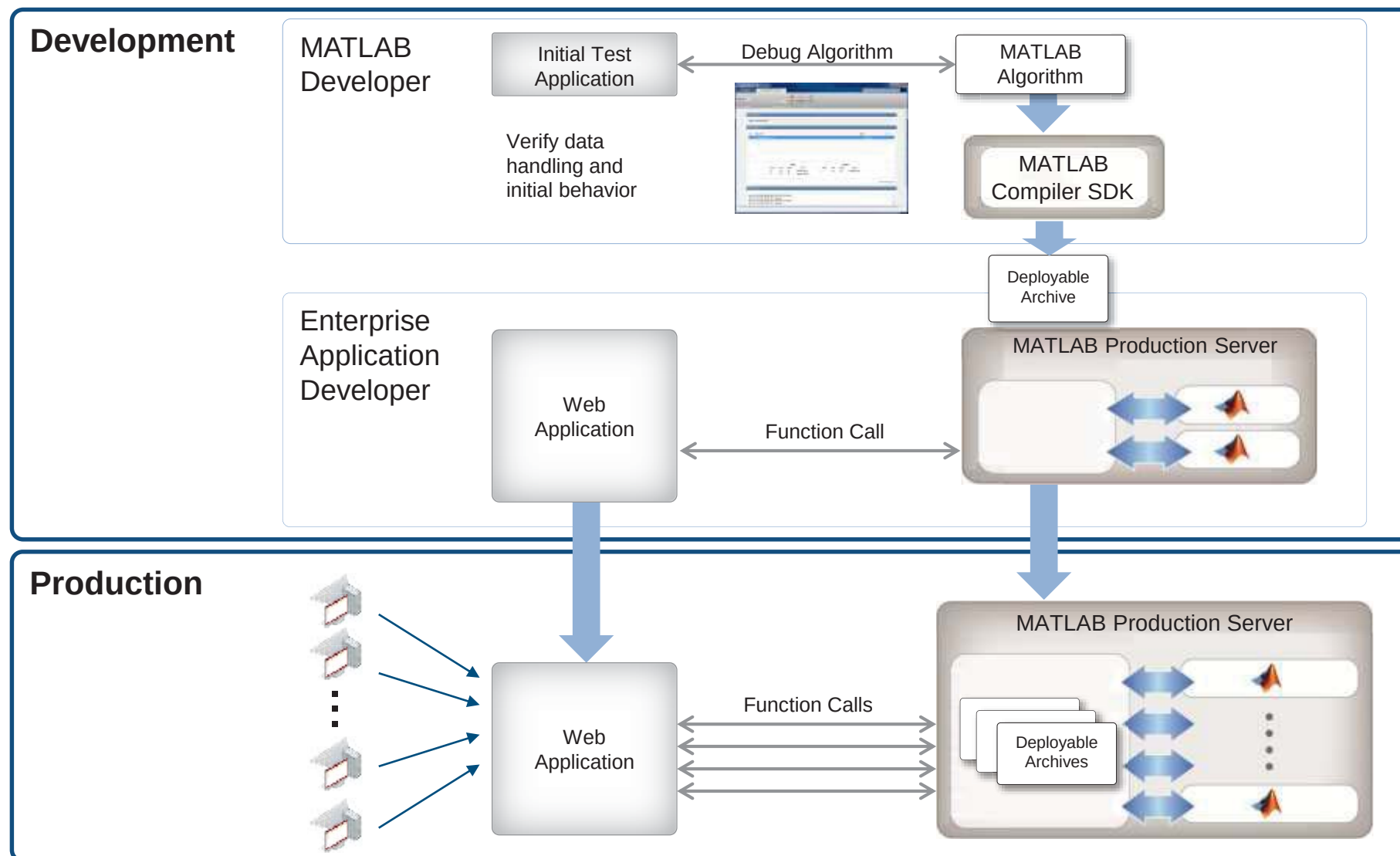
Technology Stack



Example - Integrating with IT systems



Production Deployment Workflow



Develop and Test with MATLAB Compiler SDK

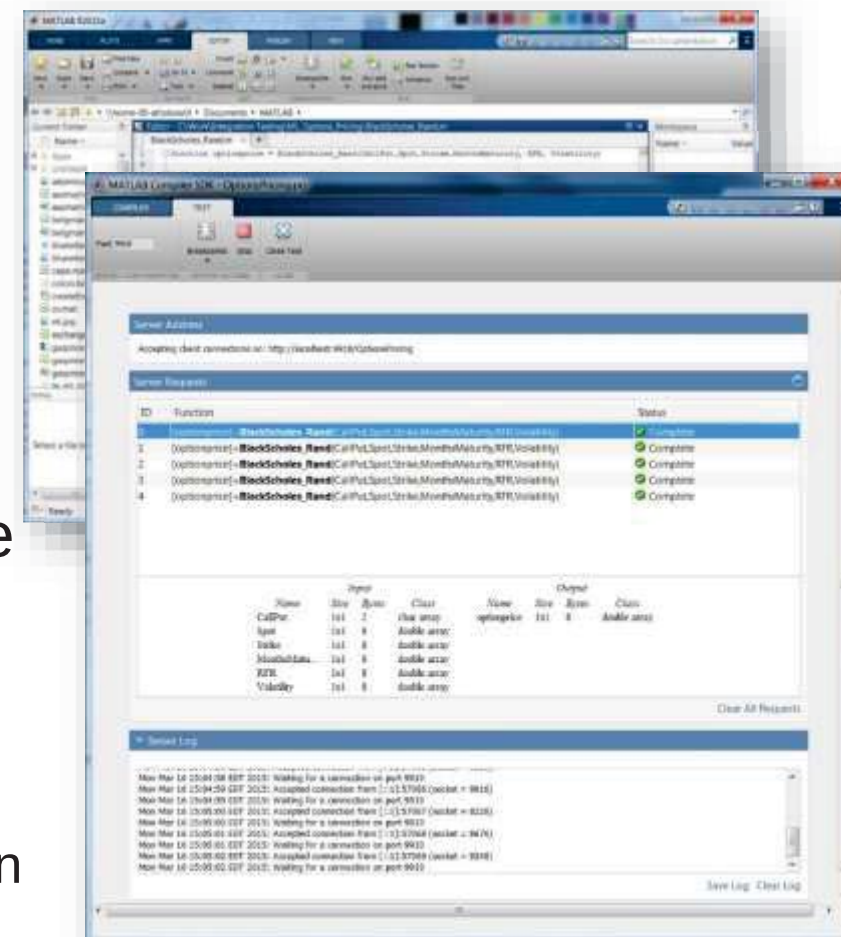
Application



HTTP

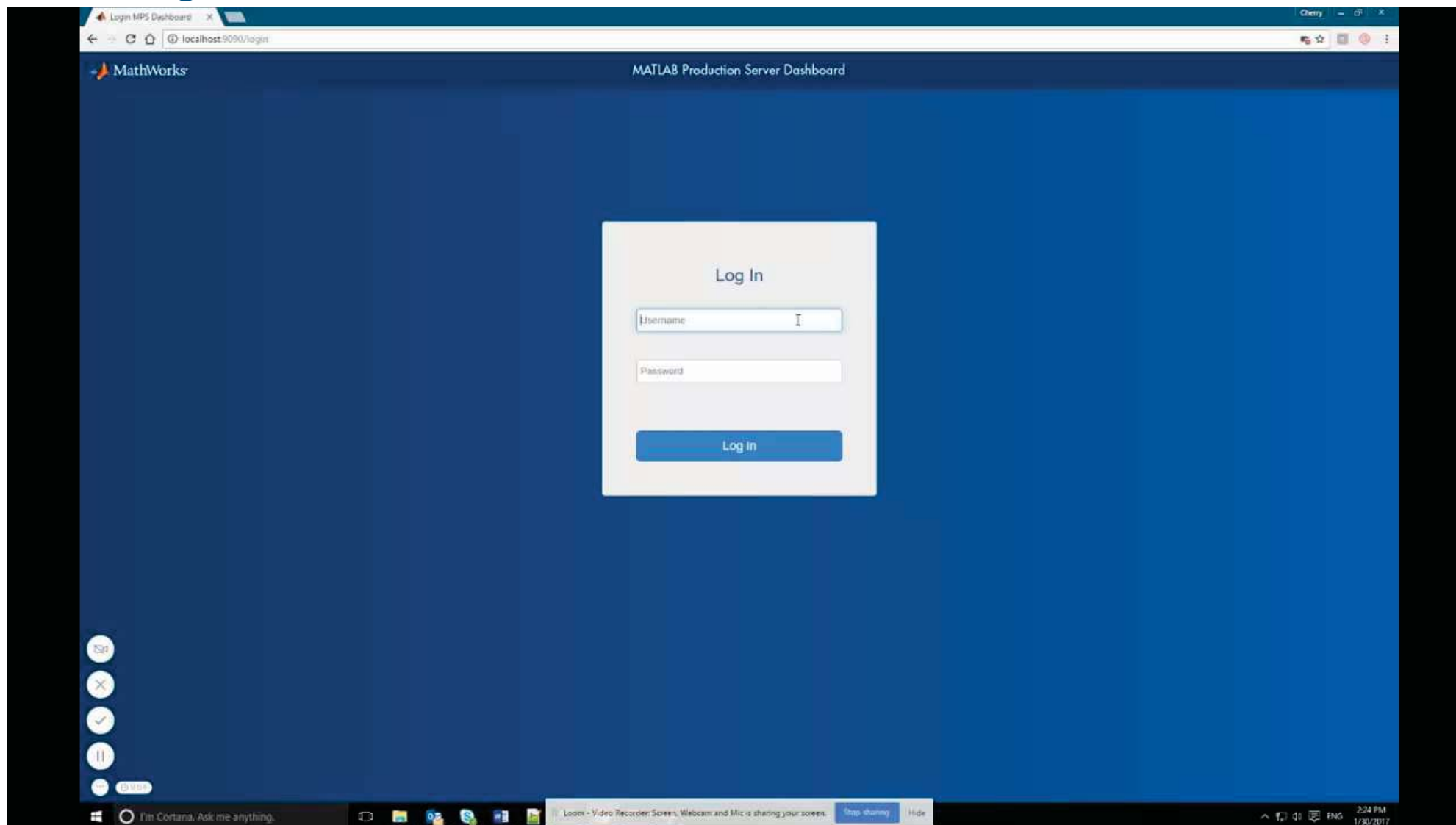


MATLAB



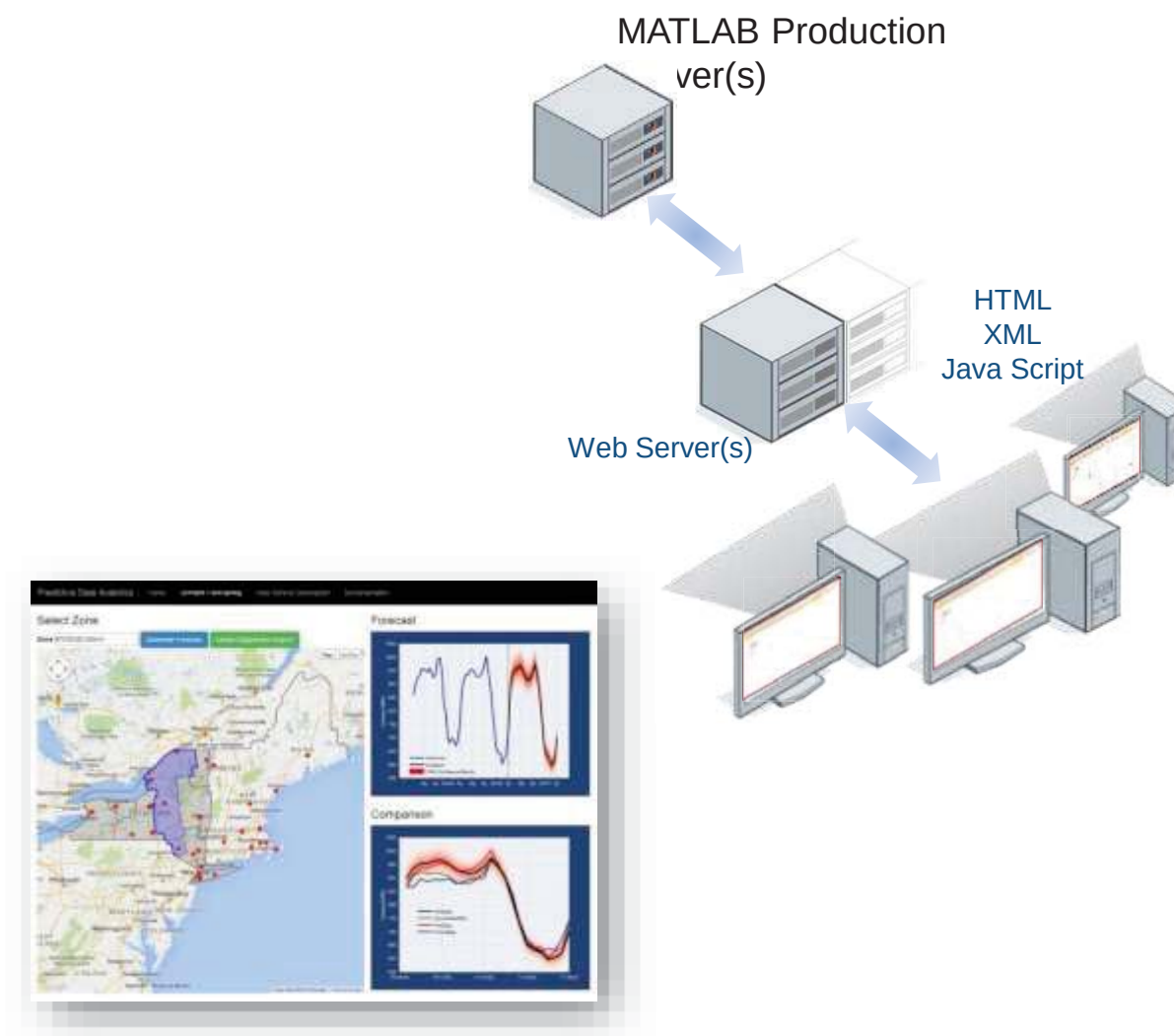
- Test environment for MATLAB Production Server
- Test and debug in MATLAB desktop
 - Details on request transactions
 - MATLAB debug and profiling with end to end testing

Web Management Dashboard – New in R2017a



Load Forecasting Demo

Energy load forecasting demo



MATLAB at Scale

MATLAB Production Server

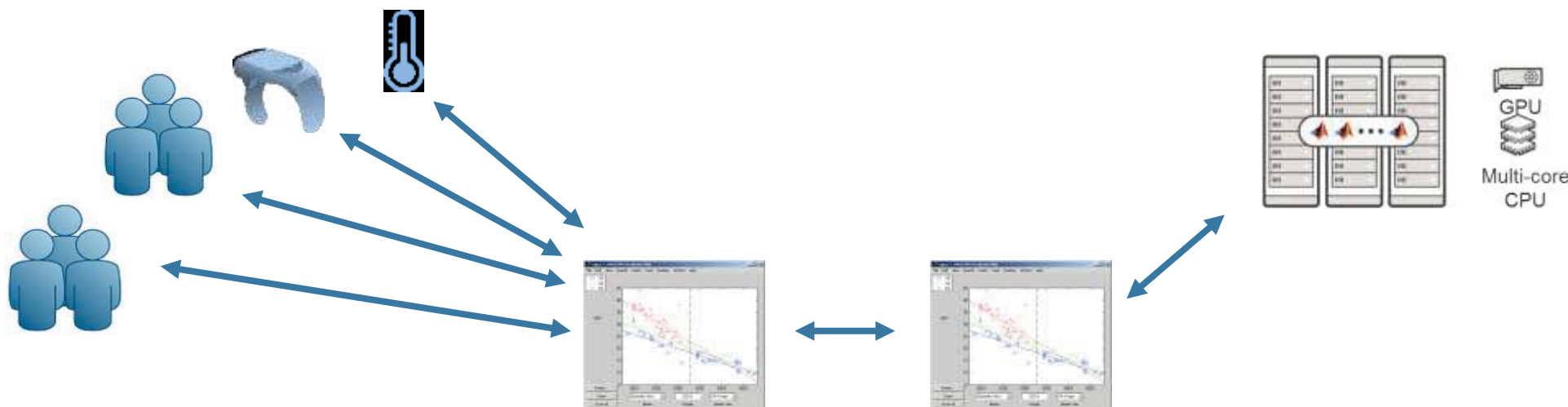
Application server for MATLAB

- Front-end scalability
- Manage large numbers of requests to run short-running deployed MATLAB programs

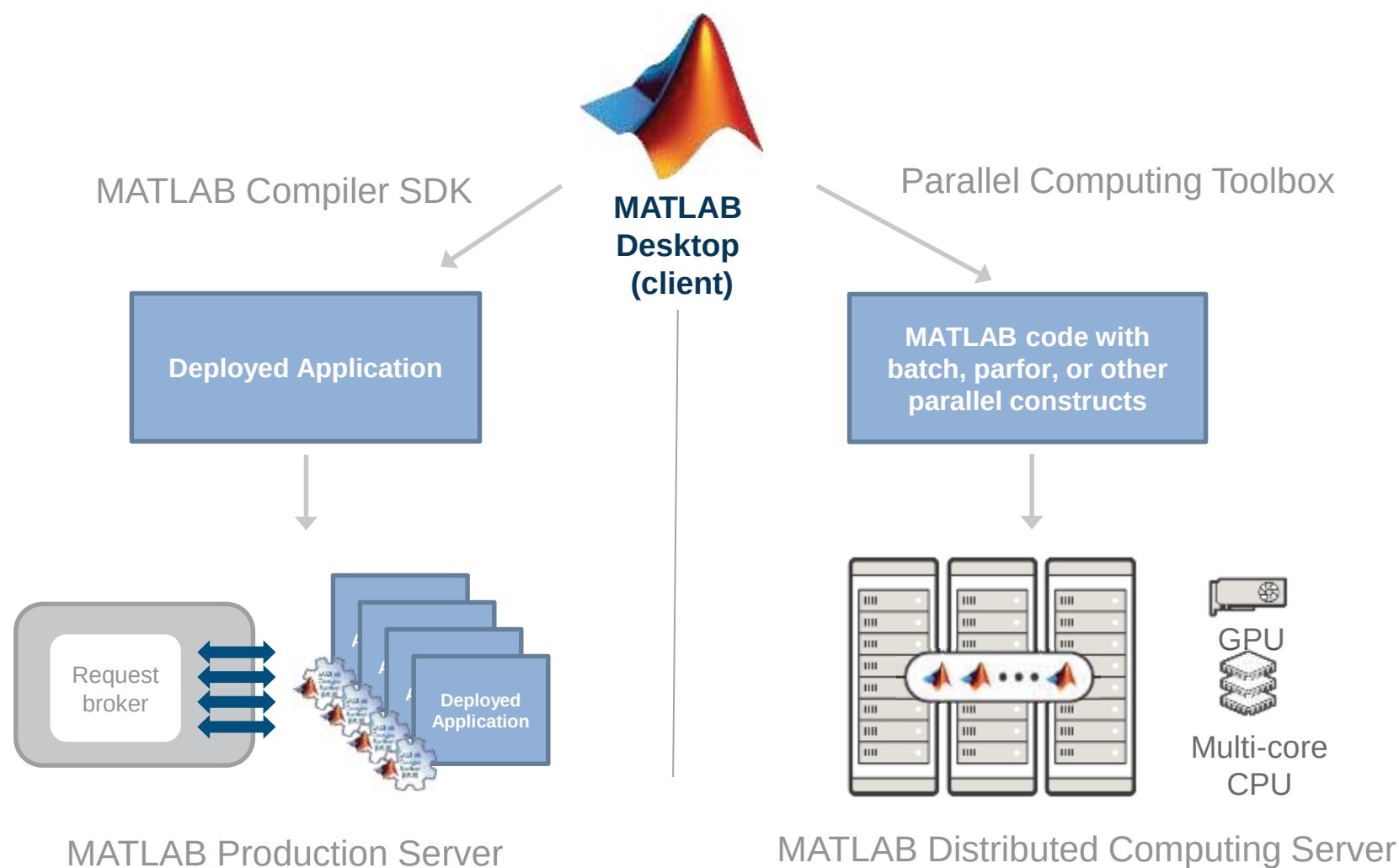
MATLAB Distributed Computing Server

Cluster framework for MATLAB/Simulink

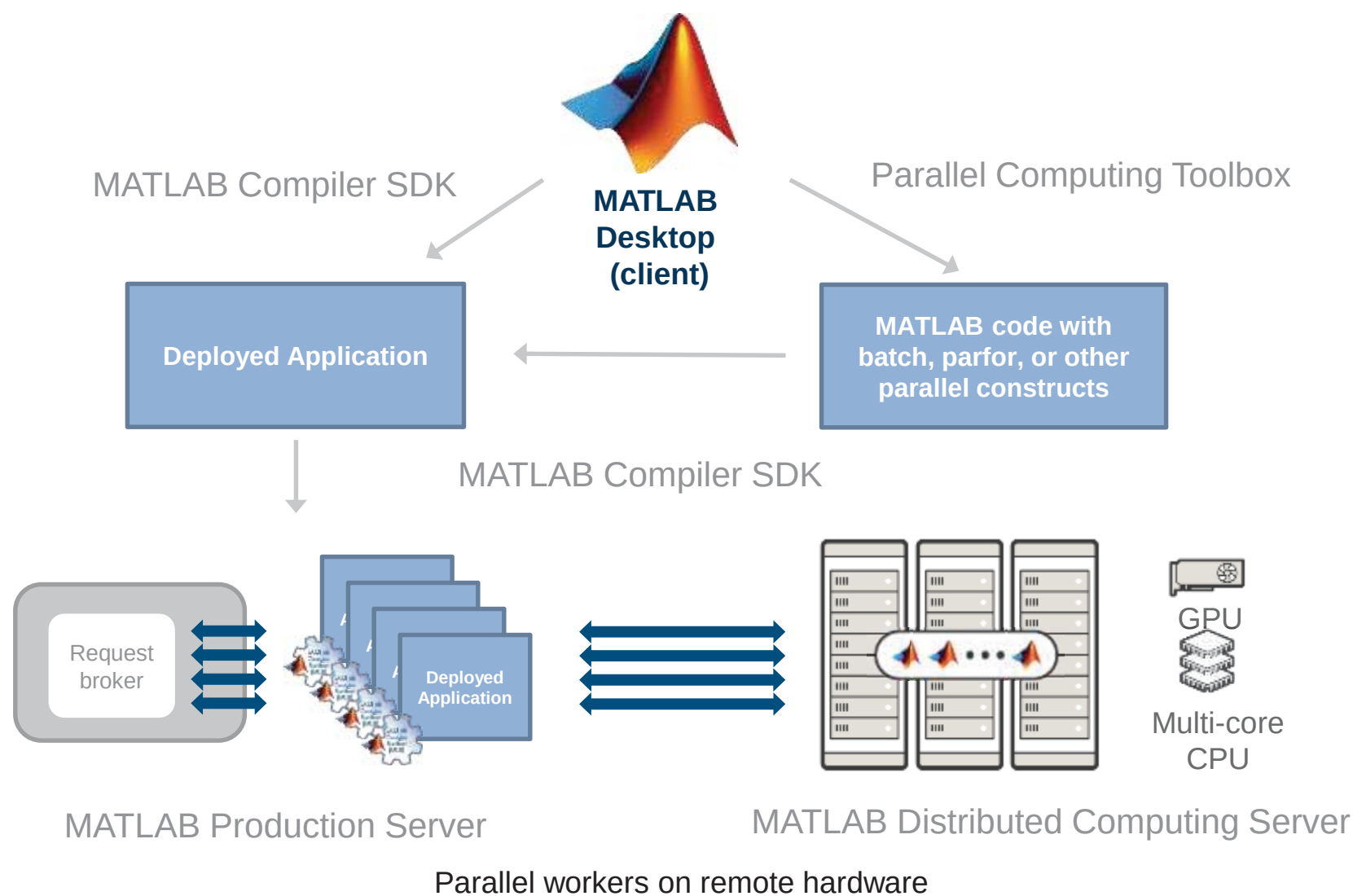
- Back-end scalability
- Speed up computationally intensive programs on computer clusters, clouds, and grids



Distinct Offerings Scale Application Access and Computation



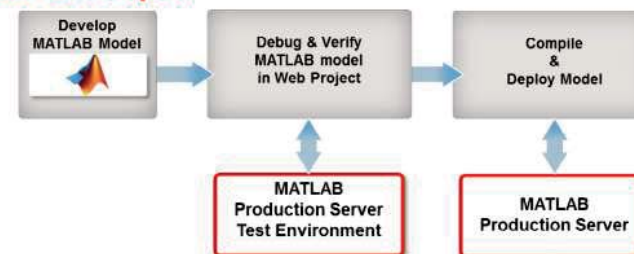
Distinct Offerings Scale Application Access and Computation



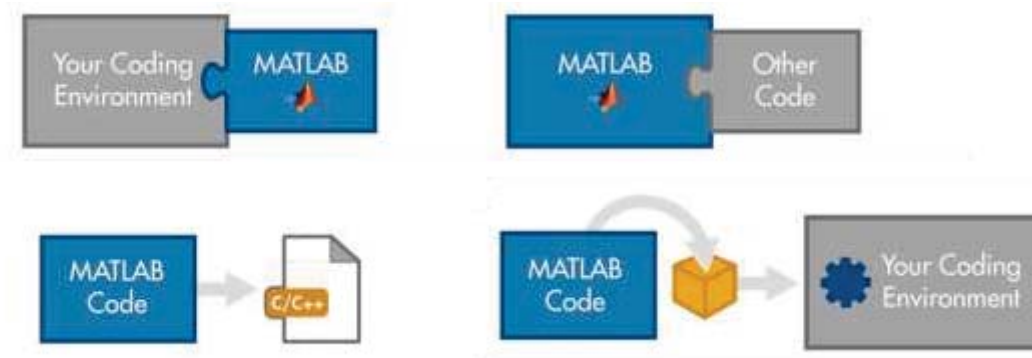
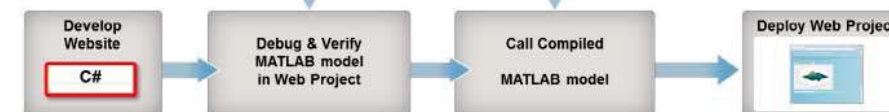
Online Resources

- Documentation – [Create and Share Toolboxes](#)
- Website – [Desktop and Web Deployment](#)
- Free White Paper – [Building a Website with MATLAB Analytics](#)
- Website – [Using MATLAB With Other Programming Languages](#)

MATLAB developer:



Web developer:

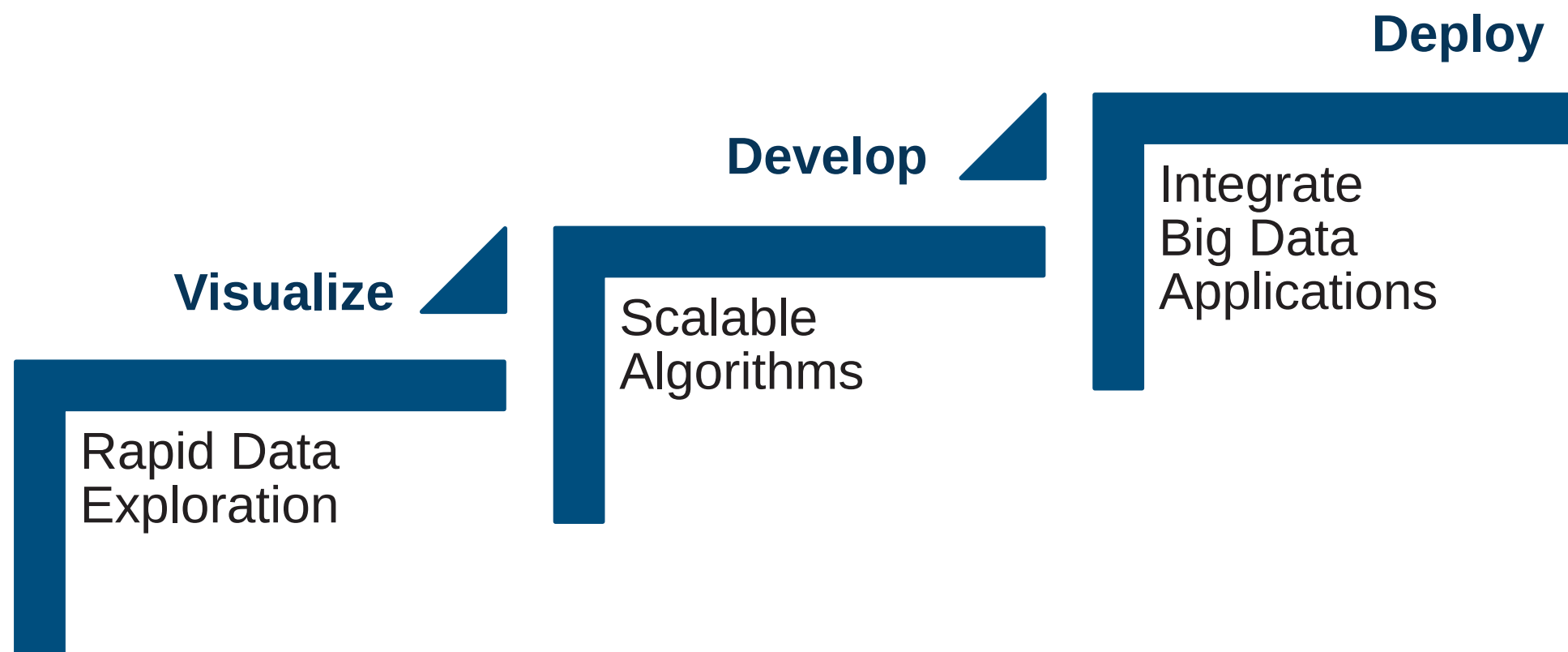


Supplemental Slides

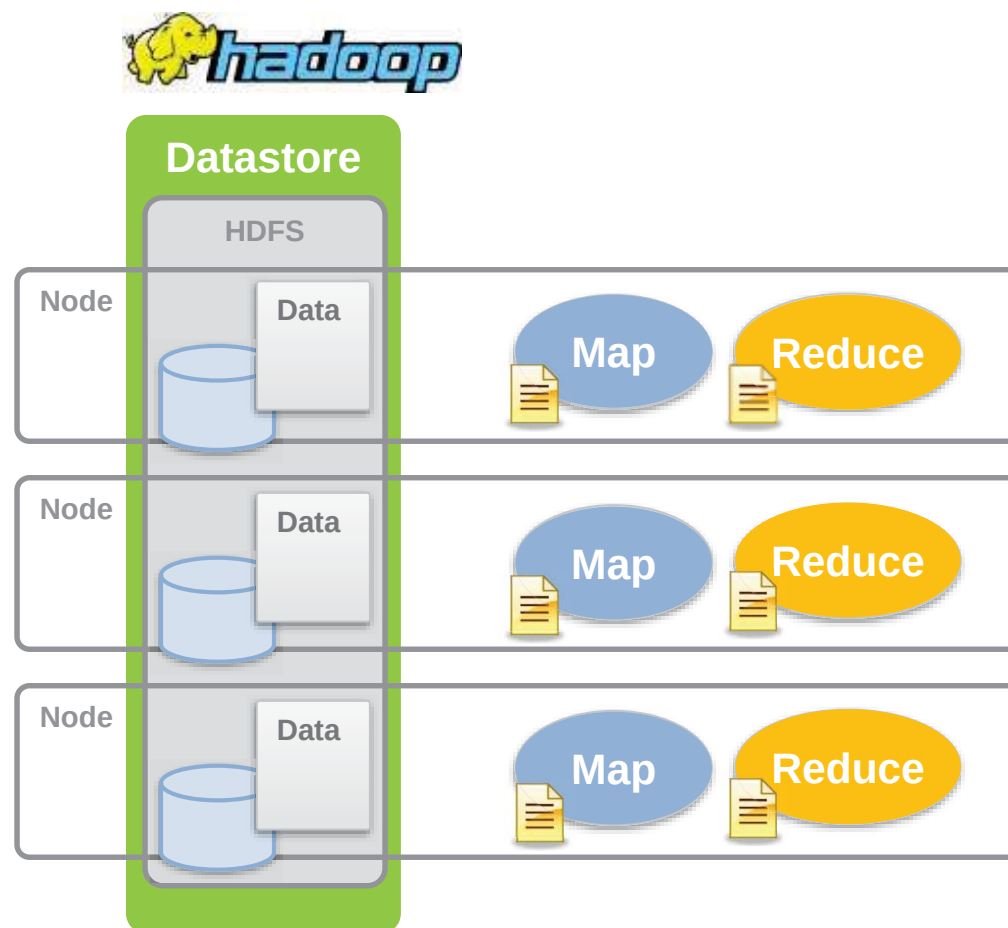
Use the following slides for more detailed discussions on various implementations using MATLAB Production Server.

Challenges of Big Data

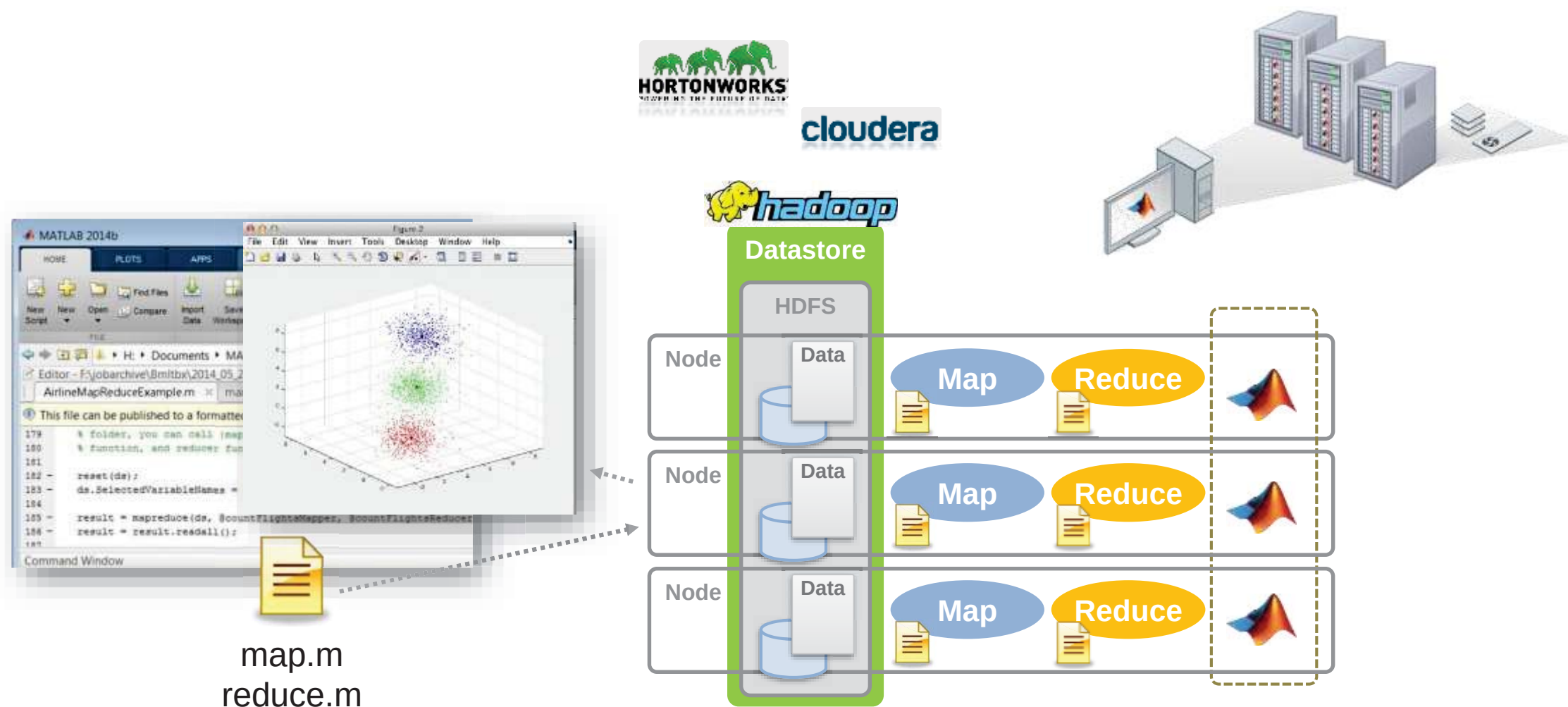
“Any collection of data sets so large and complex that it becomes difficult to process using ... traditional data processing applications.” (Wikipedia)



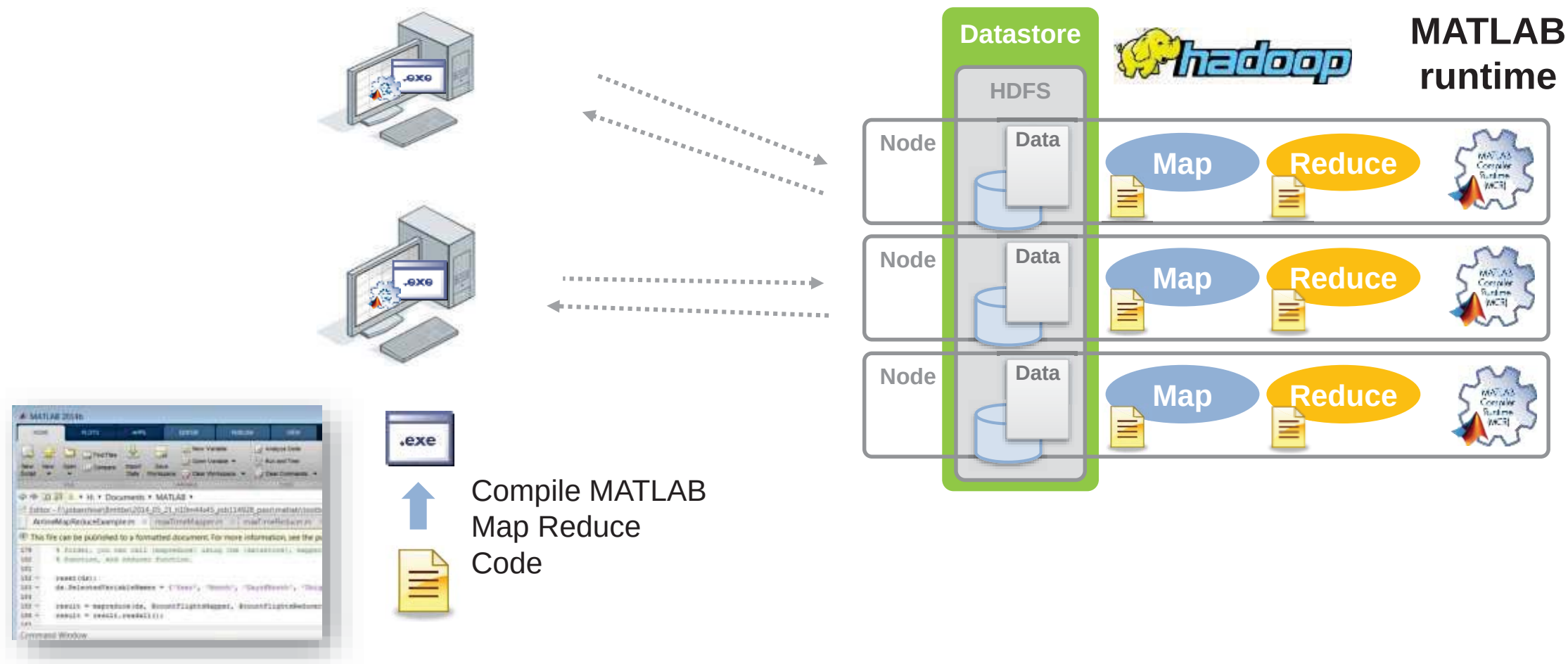
Hadoop: The Big Data Platform



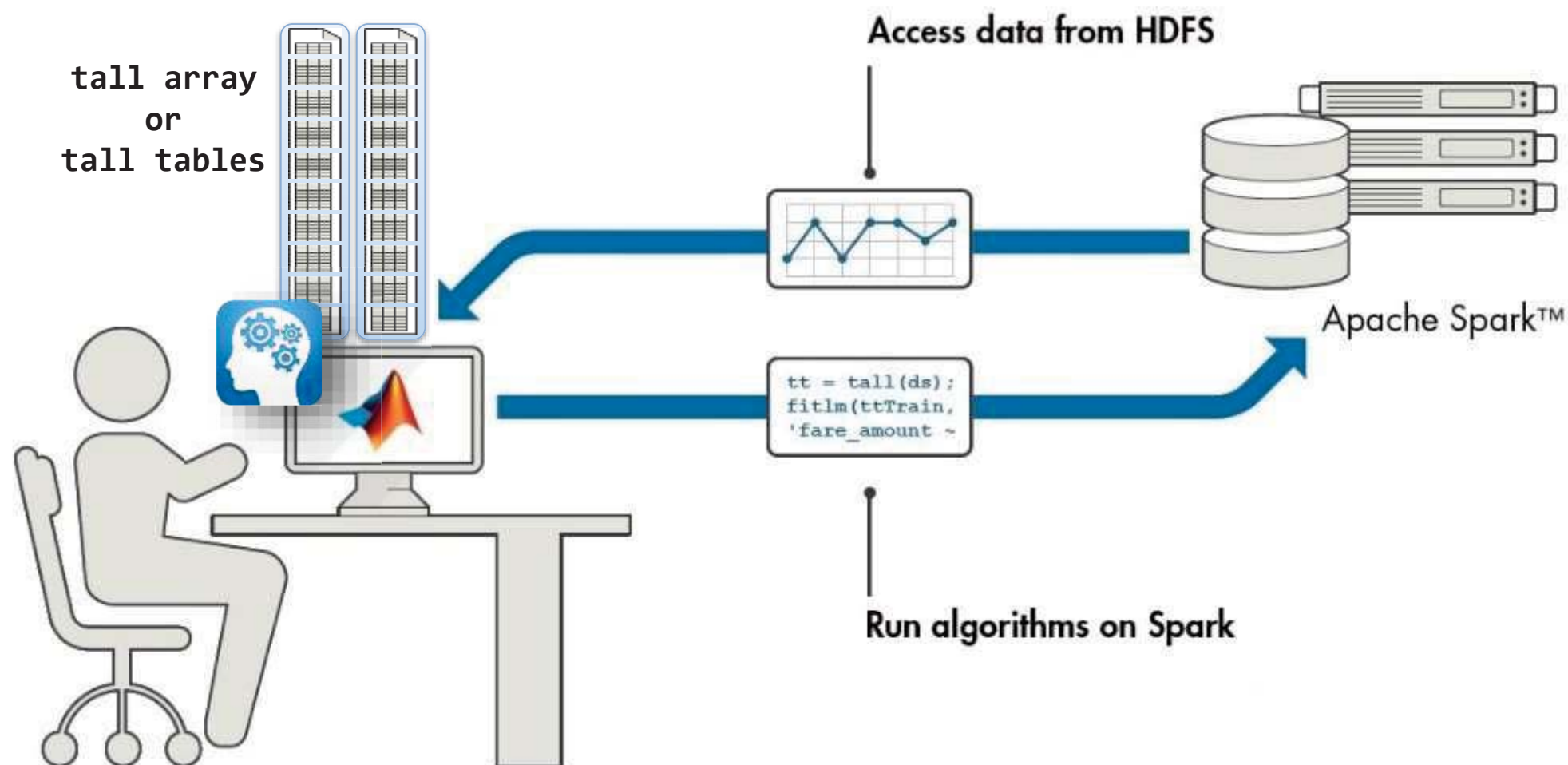
Matlab Integration with Hadoop clusters



Deploy Applications with Hadoop



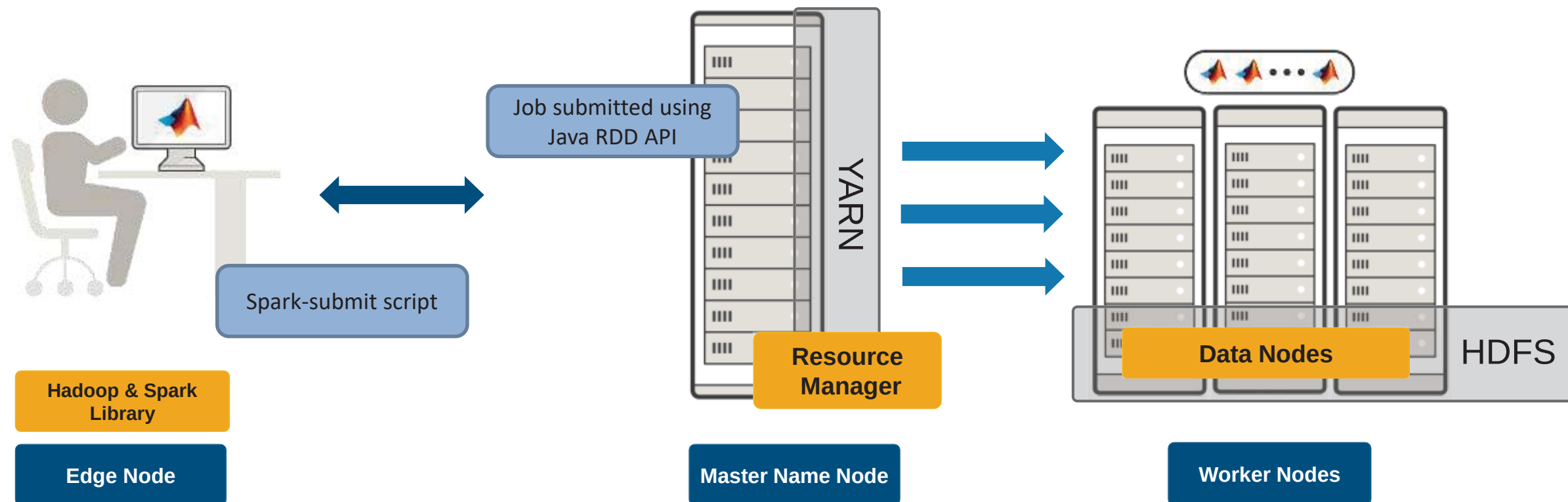
Use MATLAB with Spark on Gigabytes and Terabytes of Data



Run MATLAB scripts on SPARK & HADOOP

MATLAB workers on worker nodes in the cluster

- MDCS workers (working from MATLAB)



Example: Running on Spark enabled Hadoop

```
% Define the Execution Environment.
% Desktop
mr = mapreducer(gcp);

% Access the data.
ds = datastore('C:/datasets/taxiData/*.csv');
tt = tall(ds);
```

Desktop Code

PCT, Datastore, tall

Spark + Hadoop Code

Spark Connection

Cluster Config for Spark

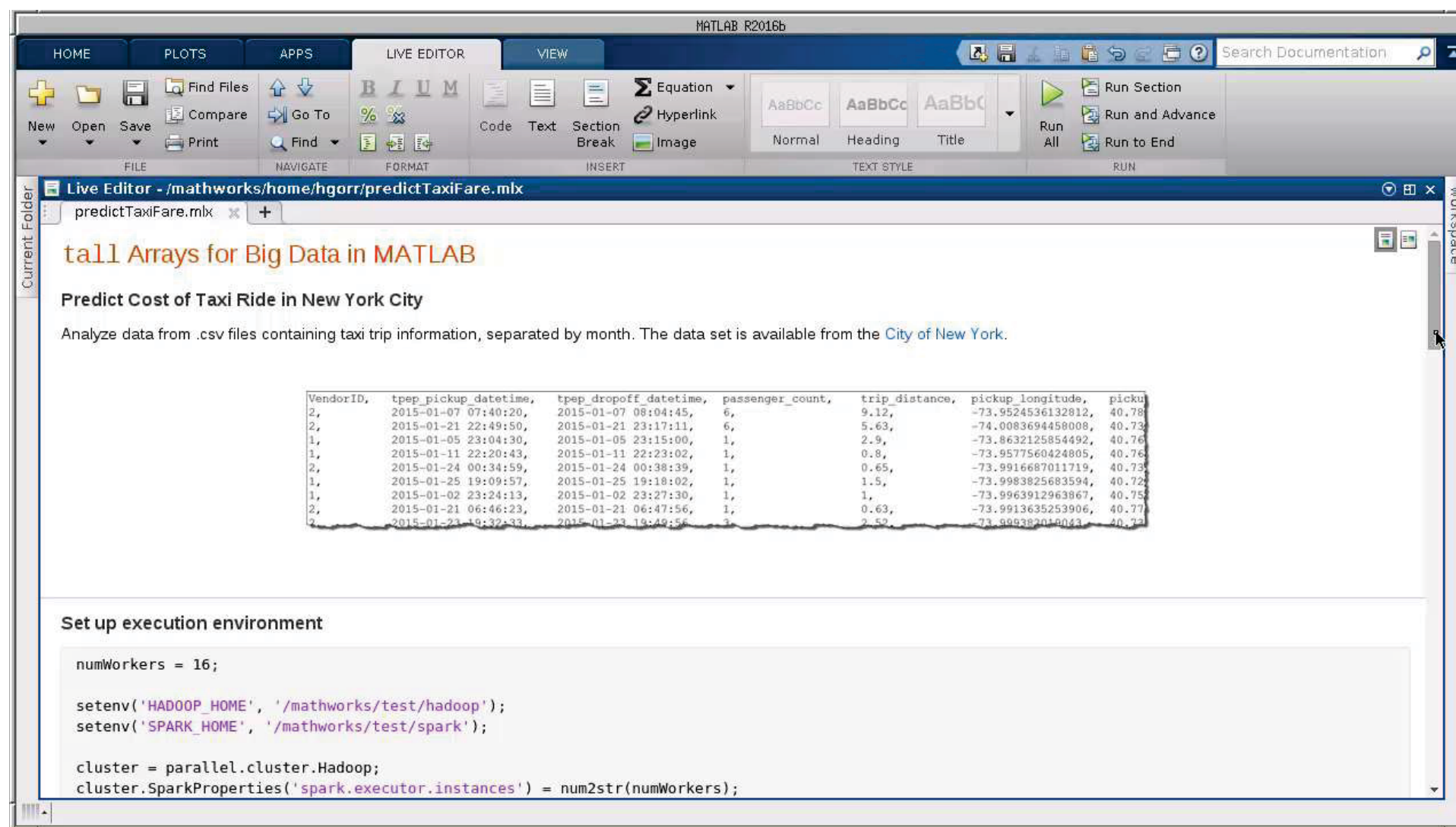
Hadoop Access

```
%% Define the Execution Environment.
% Hadoop/Spark Cluster
setenv('HADOOP_HOME', '/dev_env/cluster/hadoop');
setenv('SPARK_HOME', '/dev_env/cluster/spark');

numWorkers = 16;
cluster = parallel_cluster.Hadoop;
cluster.SparkProperties('spark.executor.instances') = num2str(numWorkers);
mr = mapreducer(cluster);

% Access the data
ds = datastore('hdfs://hadoop01:54310/datasets/taxiData/*.csv');
tt = tall(ds);
```

Example: Running on Spark and Hadoop



Live Editor - /mathworks/home/hgorr/predictTaxiFare.mlx

tall Arrays for Big Data in MATLAB

Predict Cost of Taxi Ride in New York City

Analyze data from .csv files containing taxi trip information, separated by month. The data set is available from the [City of New York](#).

VendorID,	tpep_pickup_datetime,	tpep_dropoff_datetime,	passenger_count,	trip_distance,	pickup_longitude,	pickup_latitude,
2,	2015-01-07 07:40:20,	2015-01-07 08:04:45,	6,	9.12,	-73.9524536132812,	40.7881752929688
2,	2015-01-21 22:49:50,	2015-01-21 23:17:11,	6,	5.63,	-74.0083694458008,	40.7306103515625
1,	2015-01-05 23:04:30,	2015-01-05 23:15:00,	1,	2.9,	-73.8632125854492,	40.7608752441406
1,	2015-01-11 22:20:43,	2015-01-11 22:23:02,	1,	0.8,	-73.9577560424805,	40.7608752441406
2,	2015-01-24 00:34:59,	2015-01-24 00:38:39,	1,	0.65,	-73.9916687011719,	40.7306103515625
1,	2015-01-25 19:09:57,	2015-01-25 19:18:02,	1,	1.5,	-73.9983825683594,	40.72021484375
1,	2015-01-02 23:24:13,	2015-01-02 23:27:30,	1,	1,	-73.9963912963867,	40.75021484375
2,	2015-01-21 06:46:23,	2015-01-21 06:47:56,	1,	0.63,	-73.9913635253906,	40.77021484375
2,	2015-01-23 19:32:33,	2015-01-23 19:49:56,	3,	2.52,	-73.999382010043,	40.72021484375

Set up execution environment

```
numWorkers = 16;

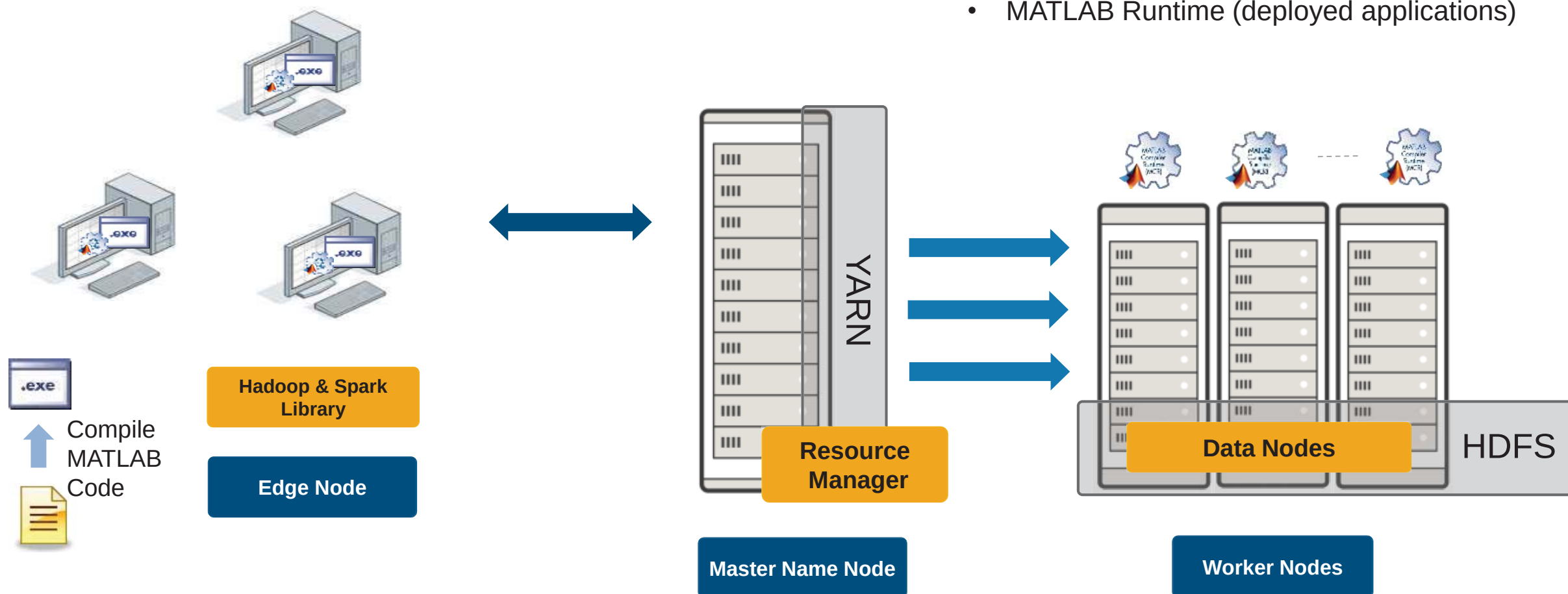
setenv('HADOOP_HOME', '/mathworks/test/hadoop');
setenv('SPARK_HOME', '/mathworks/test/spark');

cluster = parallel.cluster.Hadoop;
cluster.SparkProperties('spark.executor.instances') = num2str(numWorkers);
```

Run MATLAB scripts on SPARK & HADOOP

MATLAB workers on worker nodes in the cluster

- MATLAB Runtime (deployed applications)



Deploying Spark Applications

```
plot(result);  
legend('Actual Fare', 'Predicted Fare', 'Difference');  
savefig(f1, 'Predicted-vs-Actual.fig');
```

