

2024 MathWorks 中国汽车年会

MATLAB AI驱动应用的多场景部署 ---从嵌入式到云端

马文辉, *MathWorks* 中国




AI驱动系统开发的全流程

MATLAB&Simulink覆盖AI驱动的应用开发全流程

Data Preparation

 Data pre-processing


 Feature engineering

 Simulation-generated data

AI Modeling


 Model design and tuning


 Python interoperability

 Explain models and predictions

Simulation & Test

 Integration with complex systems

 System simulation with AI models

 System verification and validation

Deployment

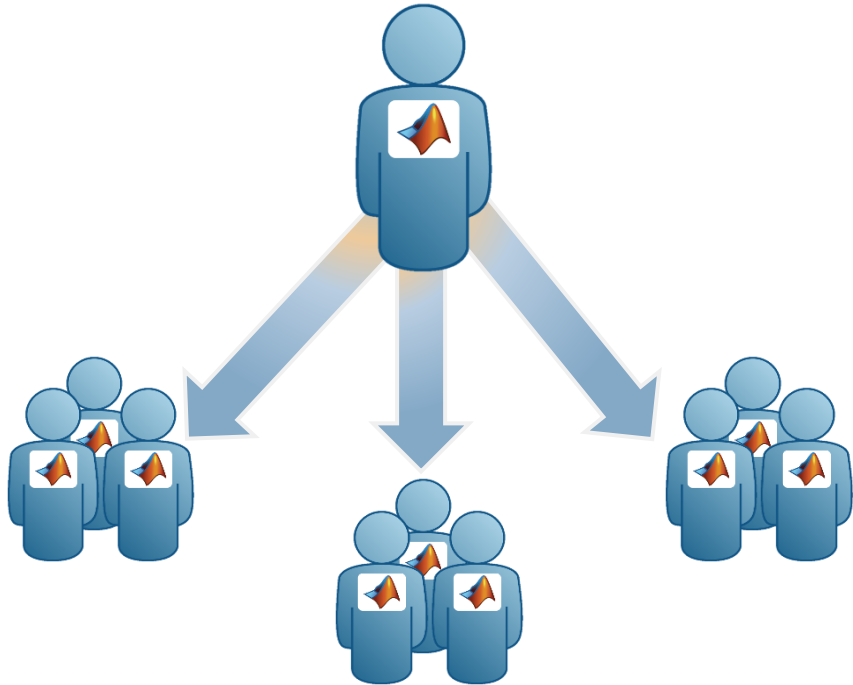
 Embedded devices

 Enterprise systems

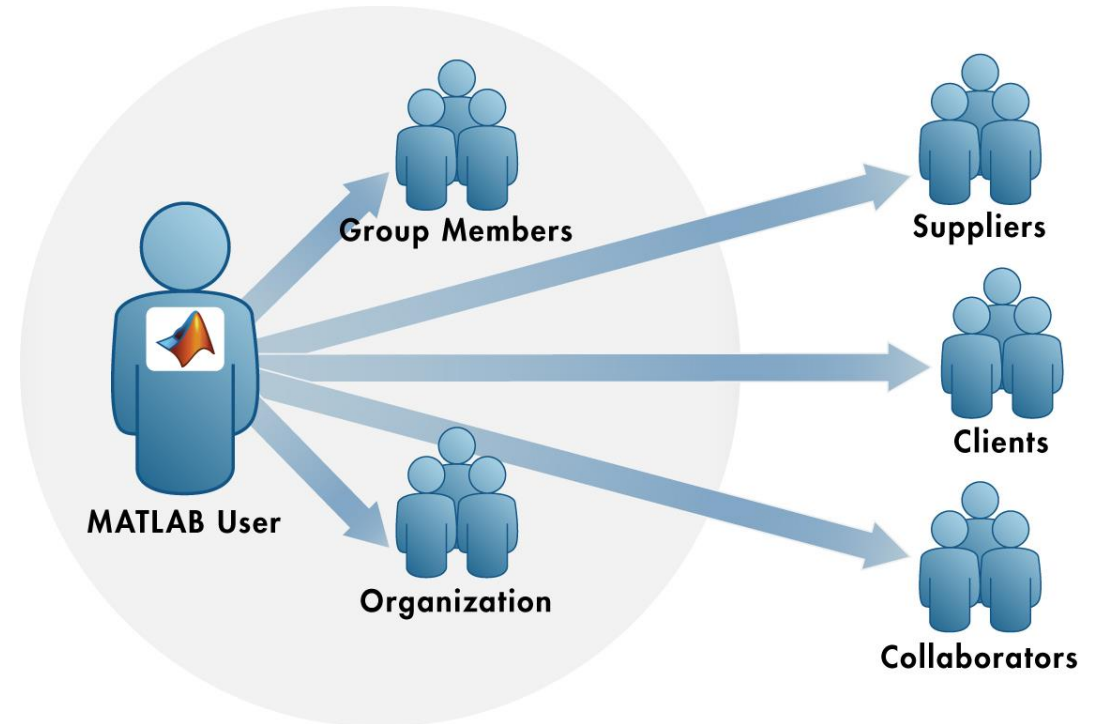
 Edge, cloud, desktop

MATLAB 应用可是实现与任何人分享（部署）

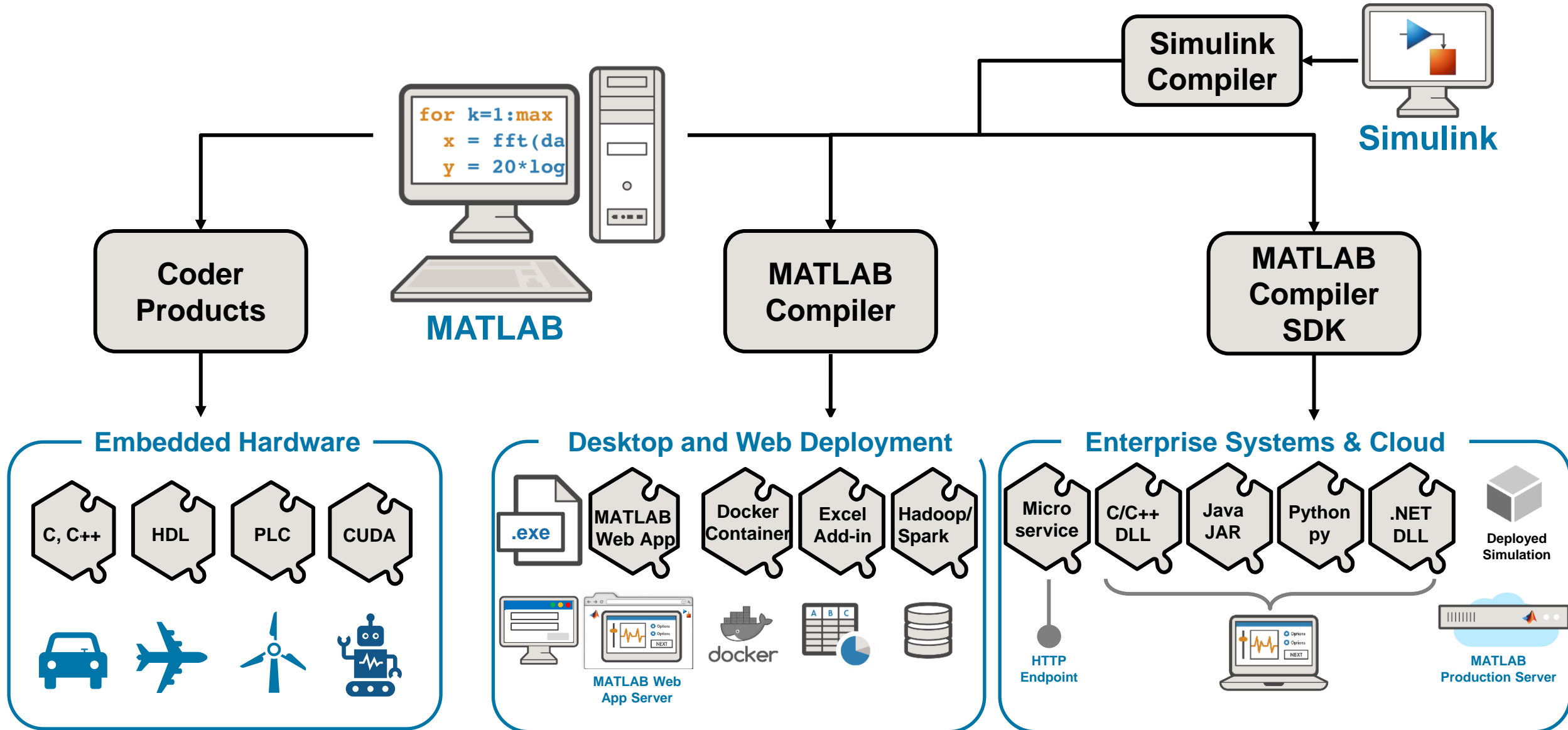
MATLAB用户（已安装MATLAB）



非MATLAB用户（未安装MATLAB）



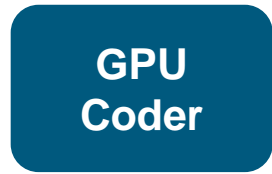
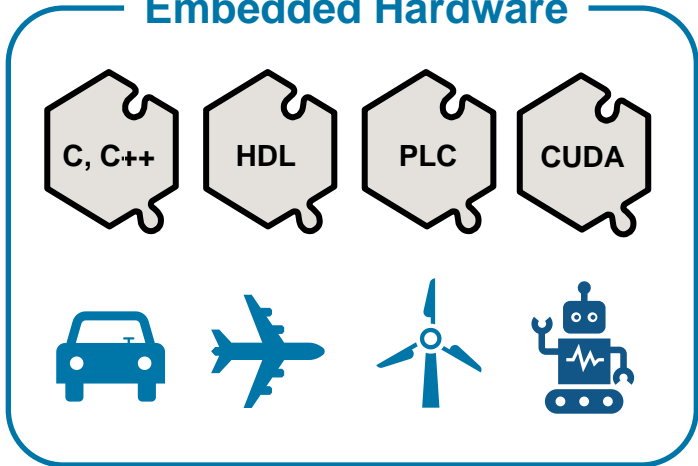
MATLAB支持应用的多场景部署



嵌入式部署



Embedded Hardware



CUDA



NVIDIA
cuDNN
& TensorRT
Libraries



自动代码生成



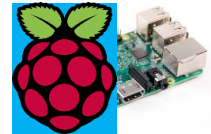
C/C++



Intel
MKL-DNN(OneDNN)
Library



ARM
Compute
Library



Verilog / VHDL



MATLAB Coder – 机器学习应用的嵌入式部署

- 故障诊断



传感器数据



ST Discovery Board Support from Embedded Coder

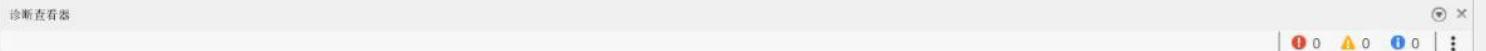
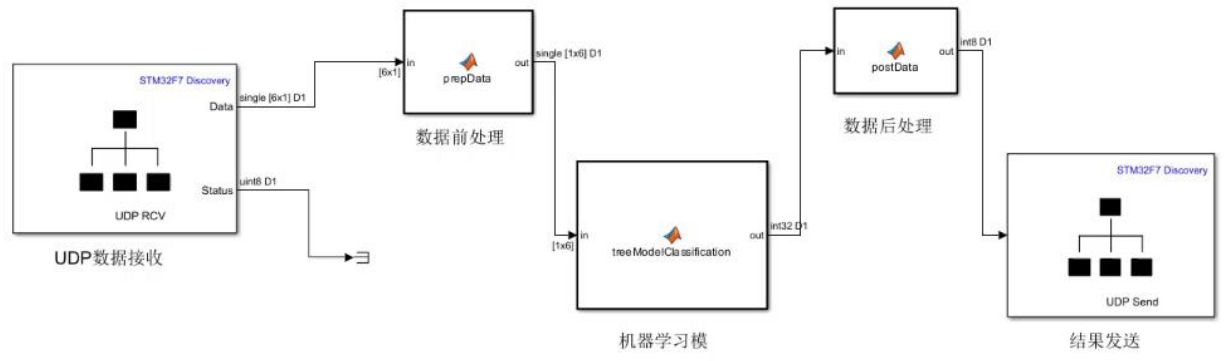
故障诊断机
器学习模型



诊断结果



基于机器学习模型实现故障诊断



就绪 95% FixedStepAuto

选择文件以查看详细信息



GPU Coder – 深度神经网络的NVIDIA CUDA代码生成

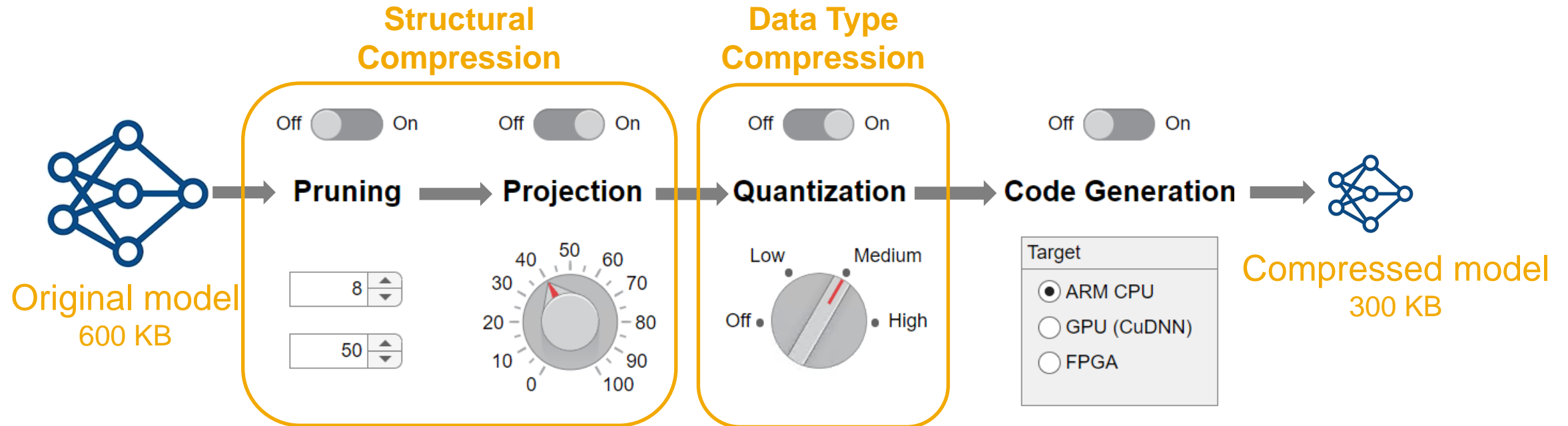
- 深度神经网络的训练
- 模型压缩
- CUDA C/C++代码生成



**NVIDIA
cuDNN
& TensorRT
Libraries**

模型压缩

降低模型在部署时对内存和计算需求



模型压缩

量化(Quantization)

- 将可学习参数从浮点数转换为整型数（int8, int16等）

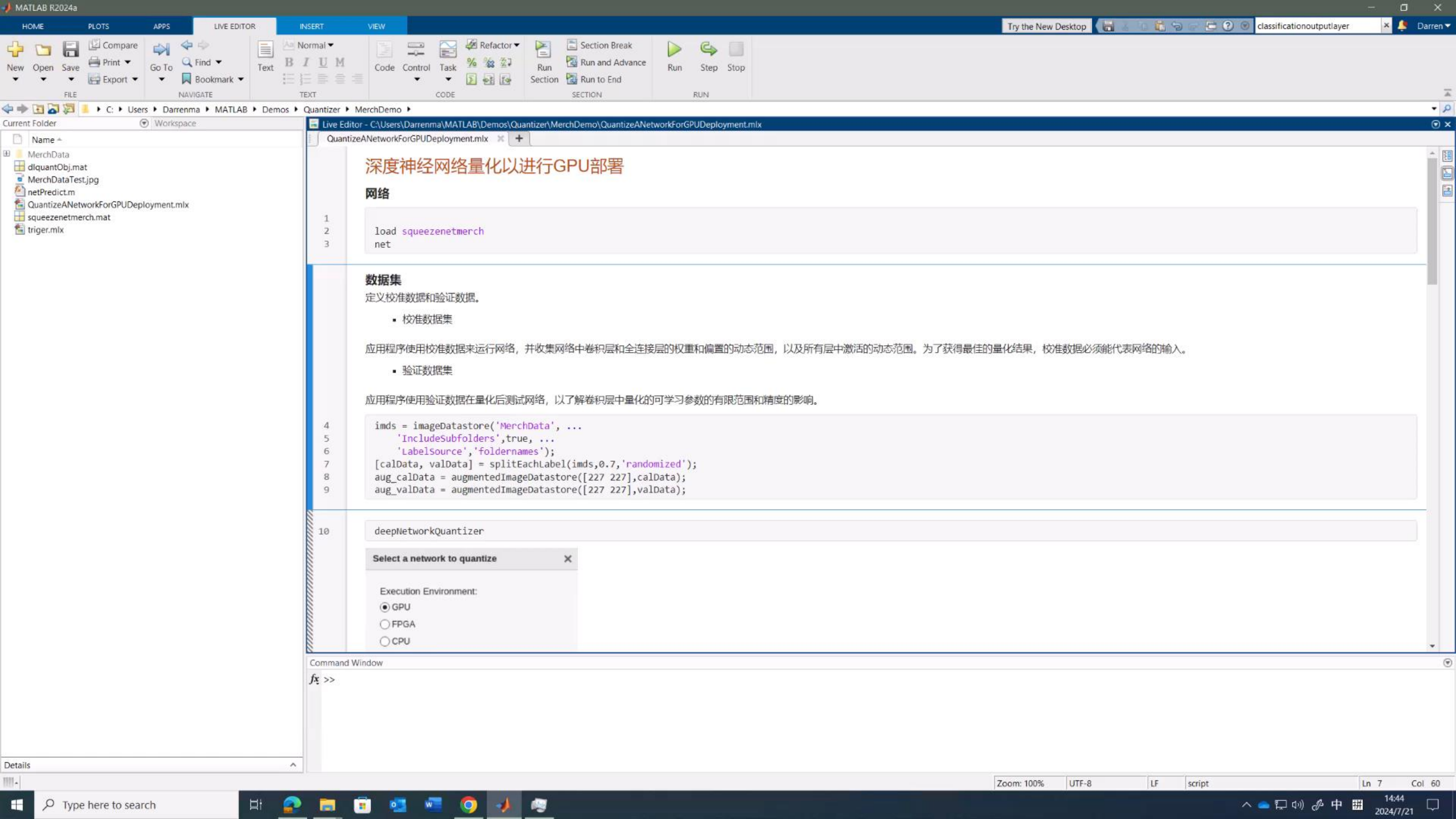
剪枝(Pruning)

- 移除卷积层中不重要的特征（或卷积核）

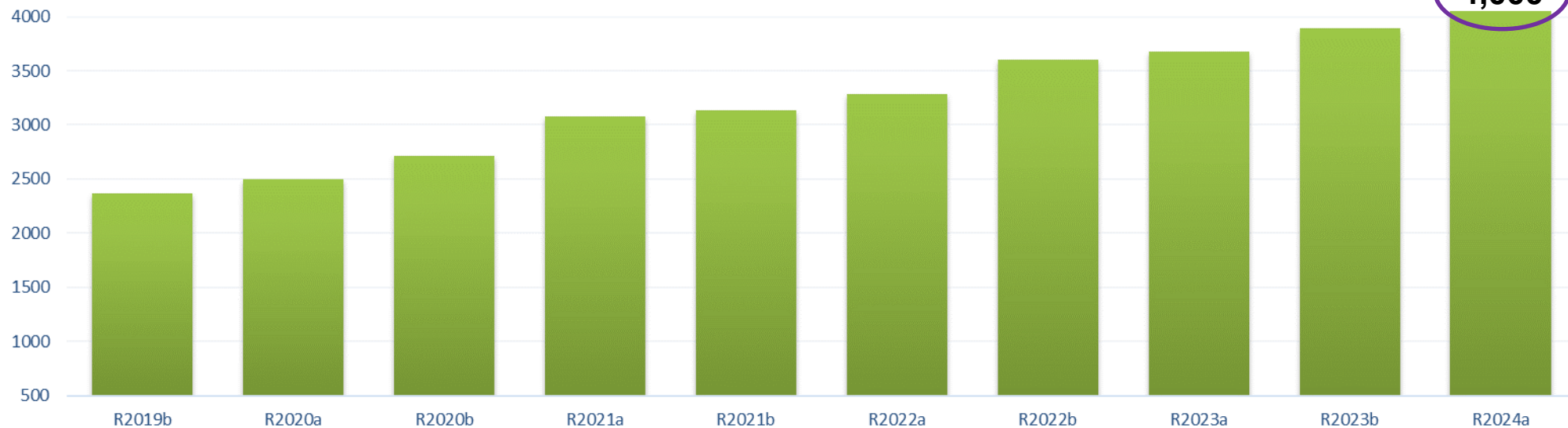
投影(Projection)

- 通过对在数据集上的层激活执行主成分分析（PCA），并对层的可学习参数应用线性投影来降低学习参数的数量

Deep Network Quantization App	R2020a
Conv2D Filter Pruning (Taylor Approximation)	R2022a
Network Projection: LSTM Layers	R2022b
Network Projection: Conv2D, FC, GRU Layers	R2023b
INT8 Arm Compute/cuDNN codegen support for dlnetworks and yoloV3/4ObjectDetectors	R2023b



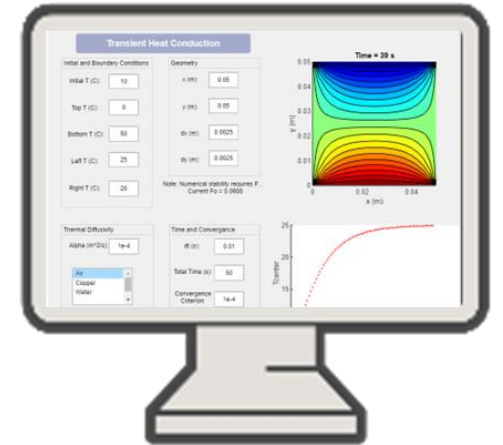
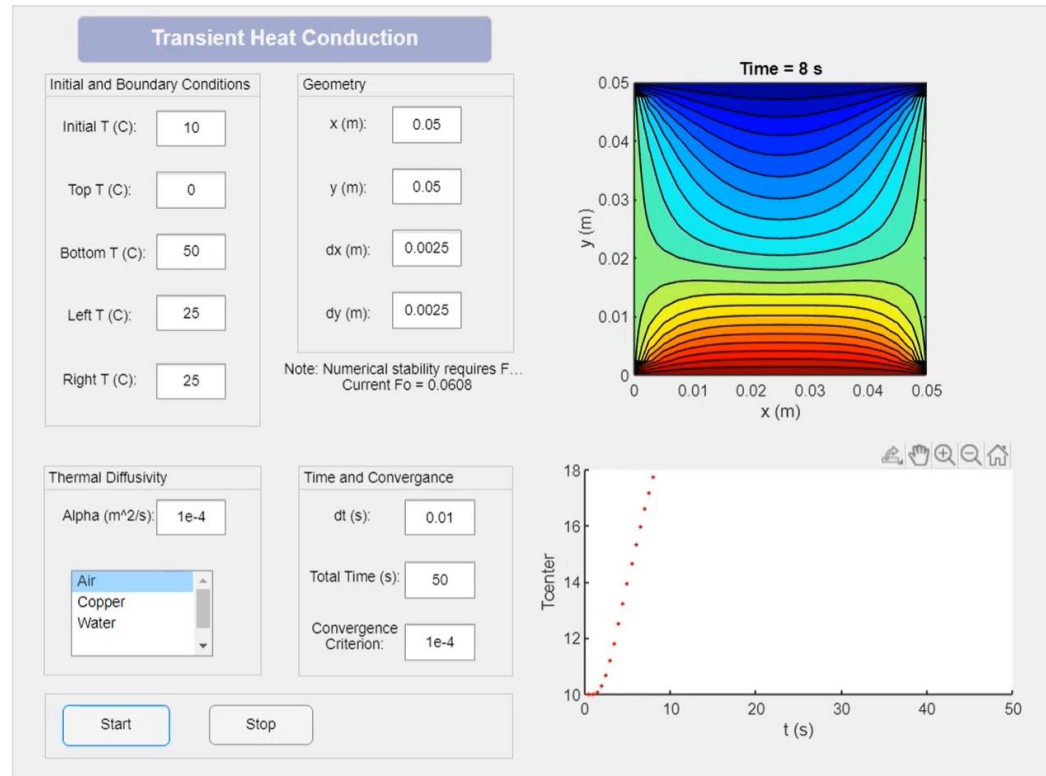
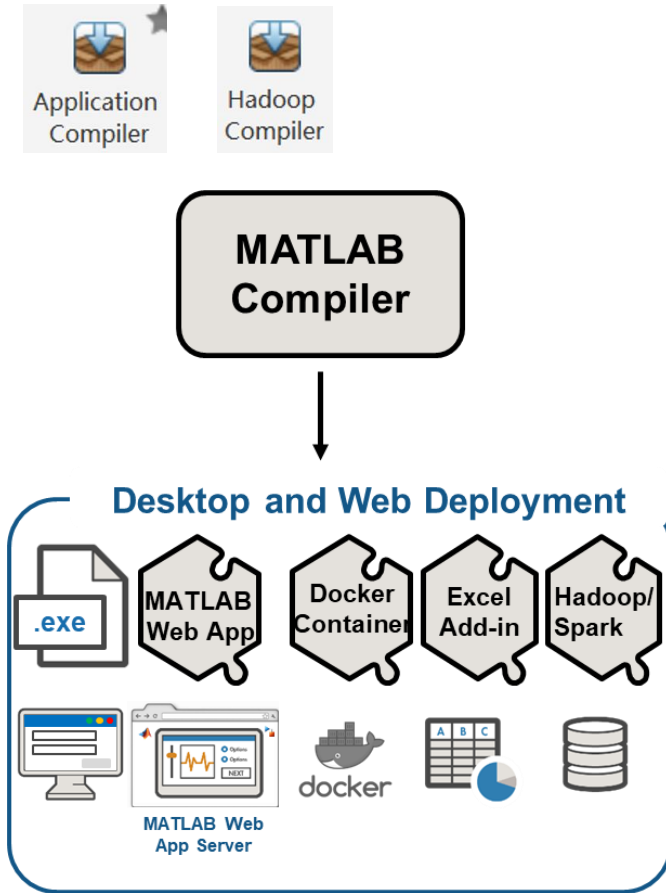
MATLAB Coder : 4,000 Functions & 39 Toolboxes Supported



- 5G Toolbox
- Aerospace Toolbox
- Antenna Toolbox
- Audio System Toolbox
- Automated Driving Toolbox
- Bluetooth Toolbox
- Communications Toolbox
- Computer Vision Toolbox
- Control System Toolbox
- Deep Learning Toolbox
- DSP System Toolbox
- Fixed-Point Designer
- Fuzzy Logic Toolbox
- Image Acquisition Toolbox
- Image Processing Toolbox
- Industrial Communication Toolbox
- Instrumental Control Toolbox
- Lidar Toolbox
- Mapping Toolbox
- Mixed-Signal Blockset
- Model Predictive Control Toolbox
- Navigation Toolbox
- Optimization Toolbox
- Phased Array System Toolbox
- Predictive Maintenance Toolbox
- Radar Toolbox
- Reinforcement Learning Toolbox
- Robotics System Toolbox
- ROS Toolbox
- Satellite Communications Toolbox
- Sensor Fusion and Tracking Toolbox
- SerDes Toolbox
- Signal Processing Toolbox
- Stats & Machine Learning Toolbox
- System Identification Toolbox
- UAV Toolbox
- Vision HDL Toolbox
- Wavelet Toolbox
- WLAN System Toolbox

边缘系统部署

- 独立应用程序

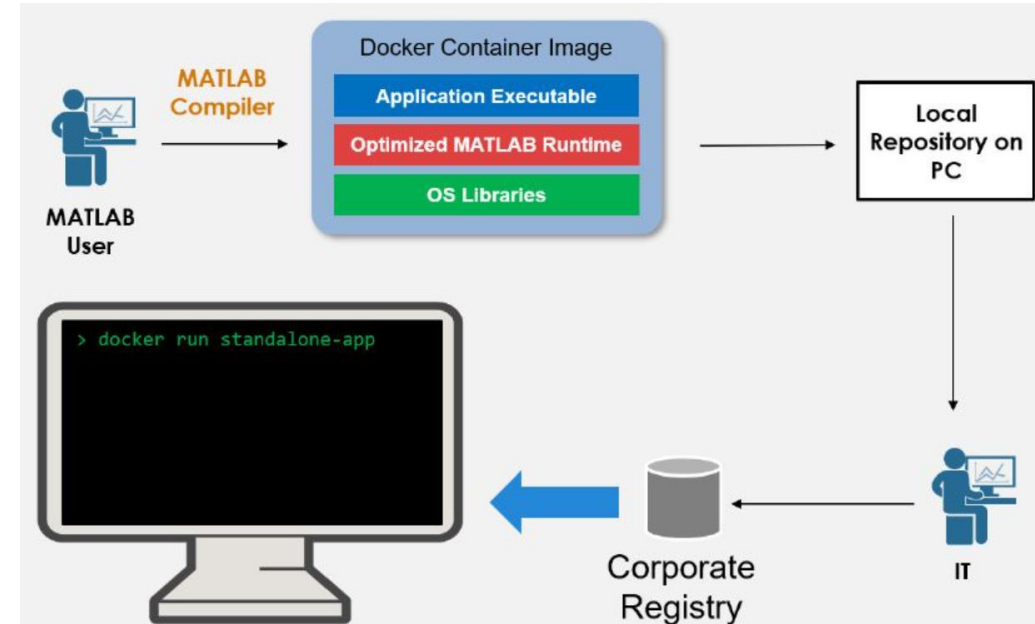
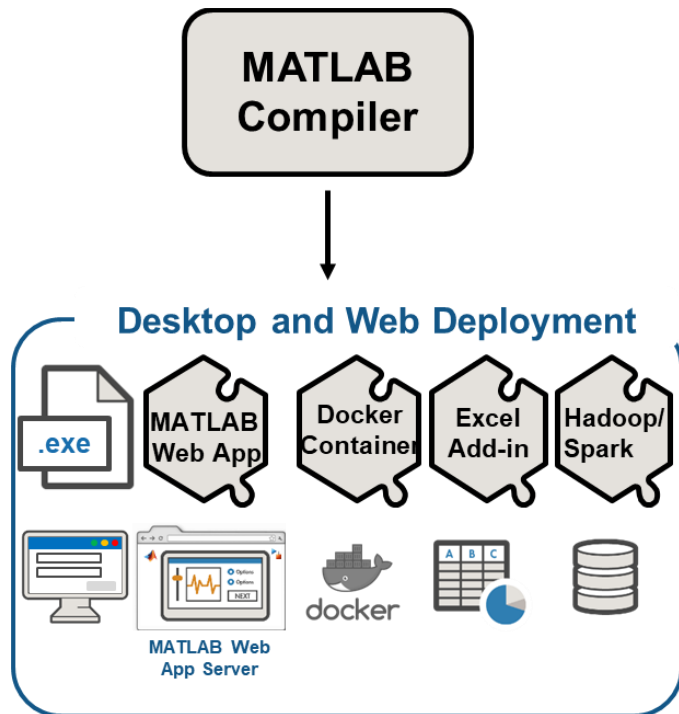


利用appdesigner开发图像化的应用程序并编译成可执行程序

边缘系统部署

- 应用程序 Docker 分发 **R2020b**

将独立应用程序作为自包含的 Docker 映像打包和分发



```
% 下载 MATLAB Runtime
compiler.runtime.download

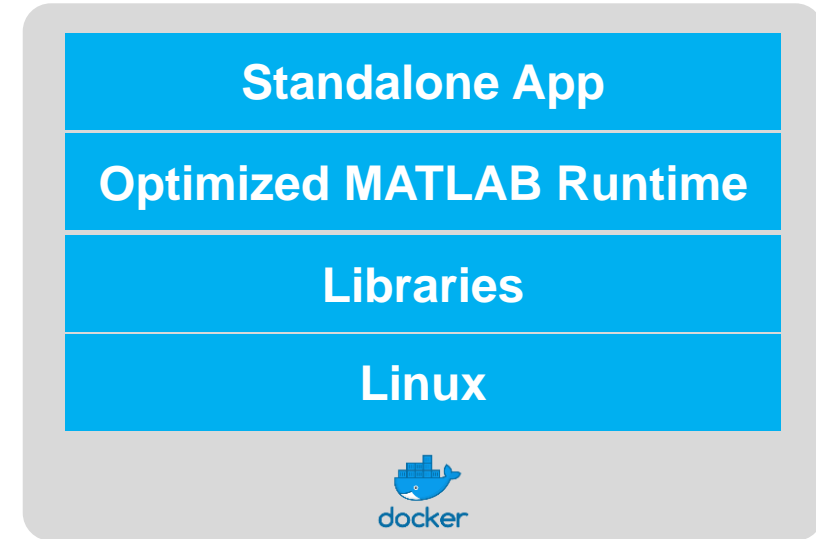
% 编译成应用程序
res = compiler.build.standaloneApplication('magicDemo.m', 'TreatInputsAsNumeric', true);

% 生成docker image
opts = compiler.package.DockerOptions(res, 'ImageName', 'magicDemo-standalone-app');
compiler.package.docker(res, 'Options', opts)
```

将MATLAB应用编译成Docker Container Images

Package: Linux-only

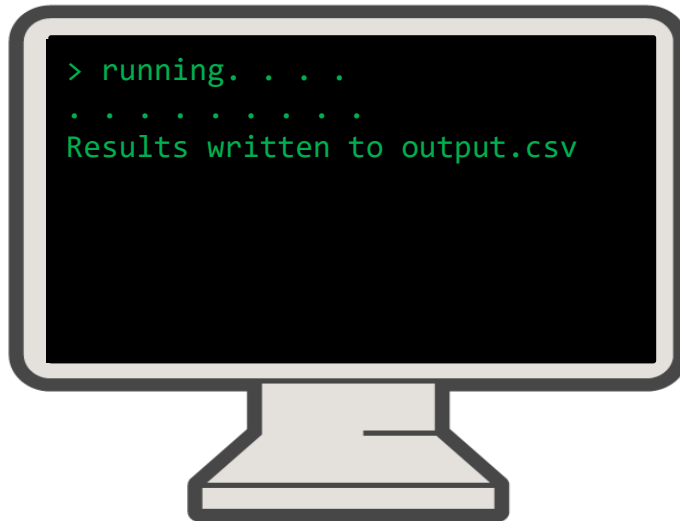
```
> r = compiler.build.StandaloneApplication ('file.m',..)
> opts = compiler.package.DockerOptions(..)
> > compiler.package.docker(r,"Options",opt)
```



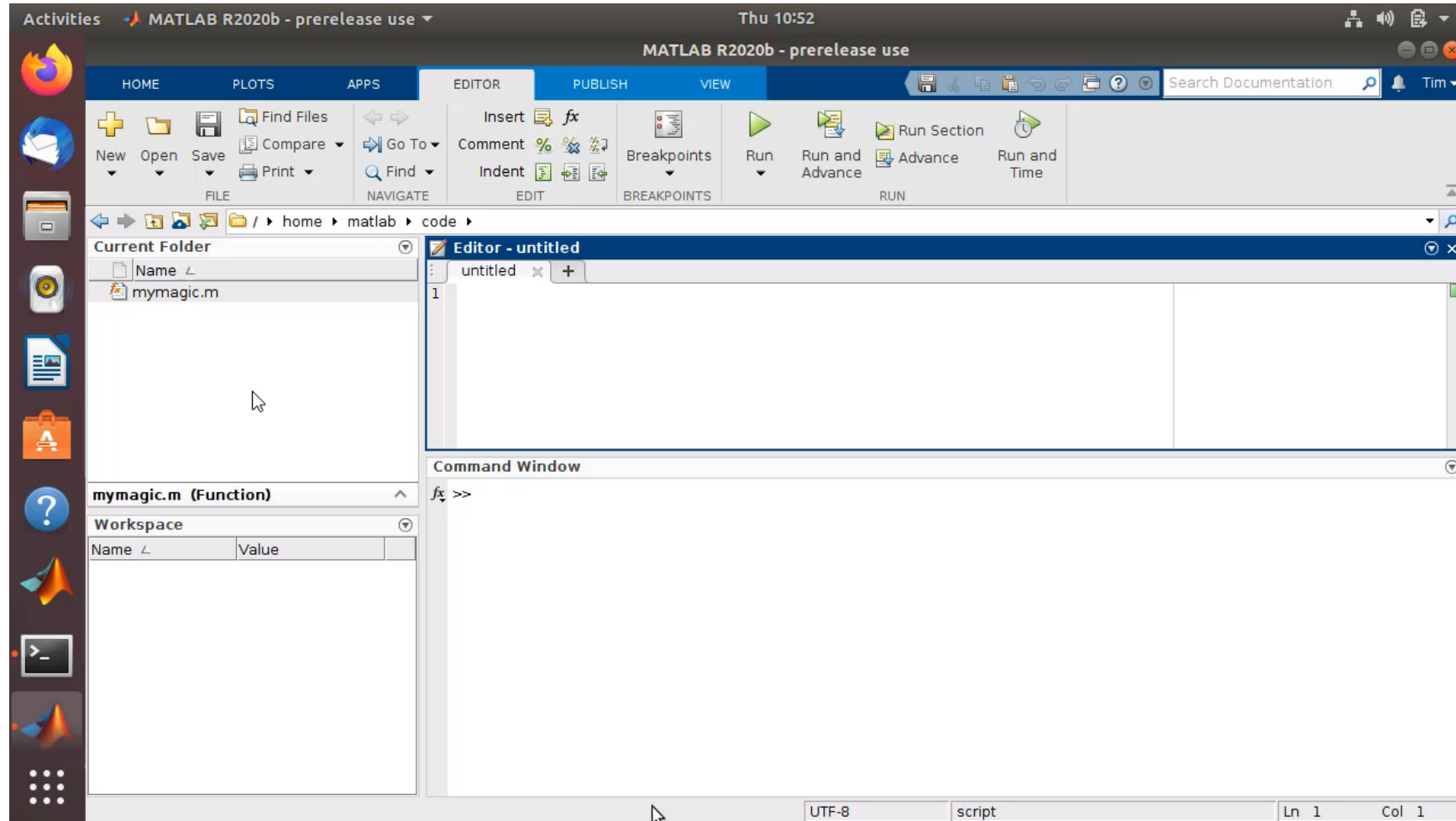
In your local Docker repository



Docker push to remote repository



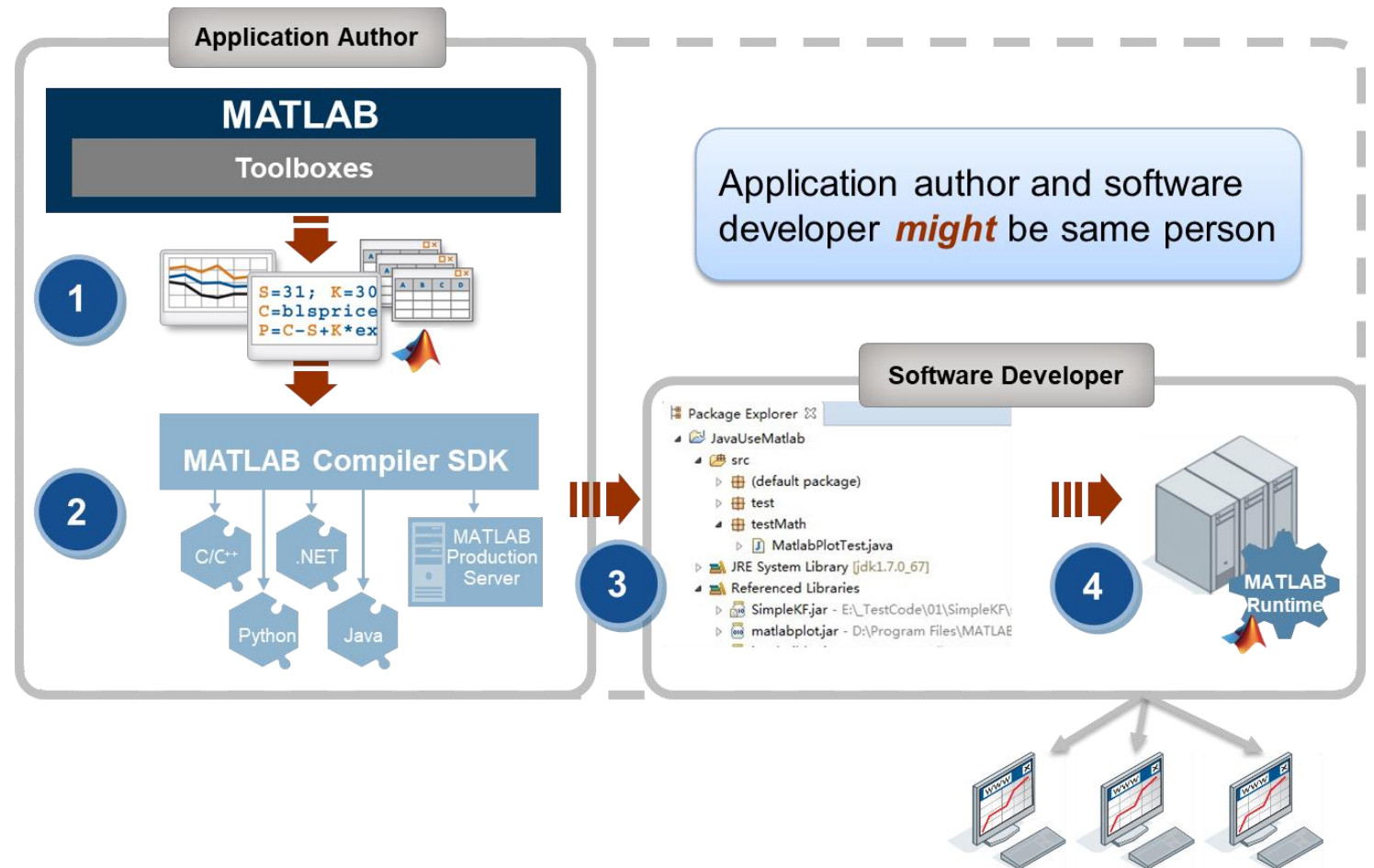
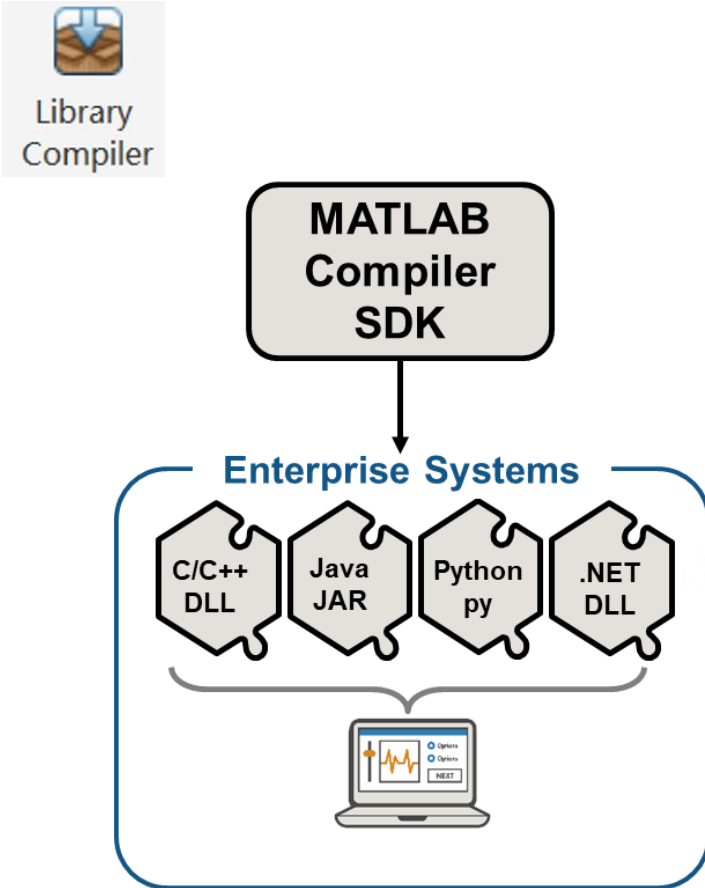
将MATLAB应用编译成Docker Container Images



R2020b

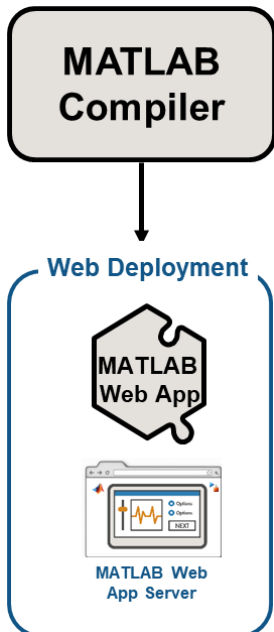
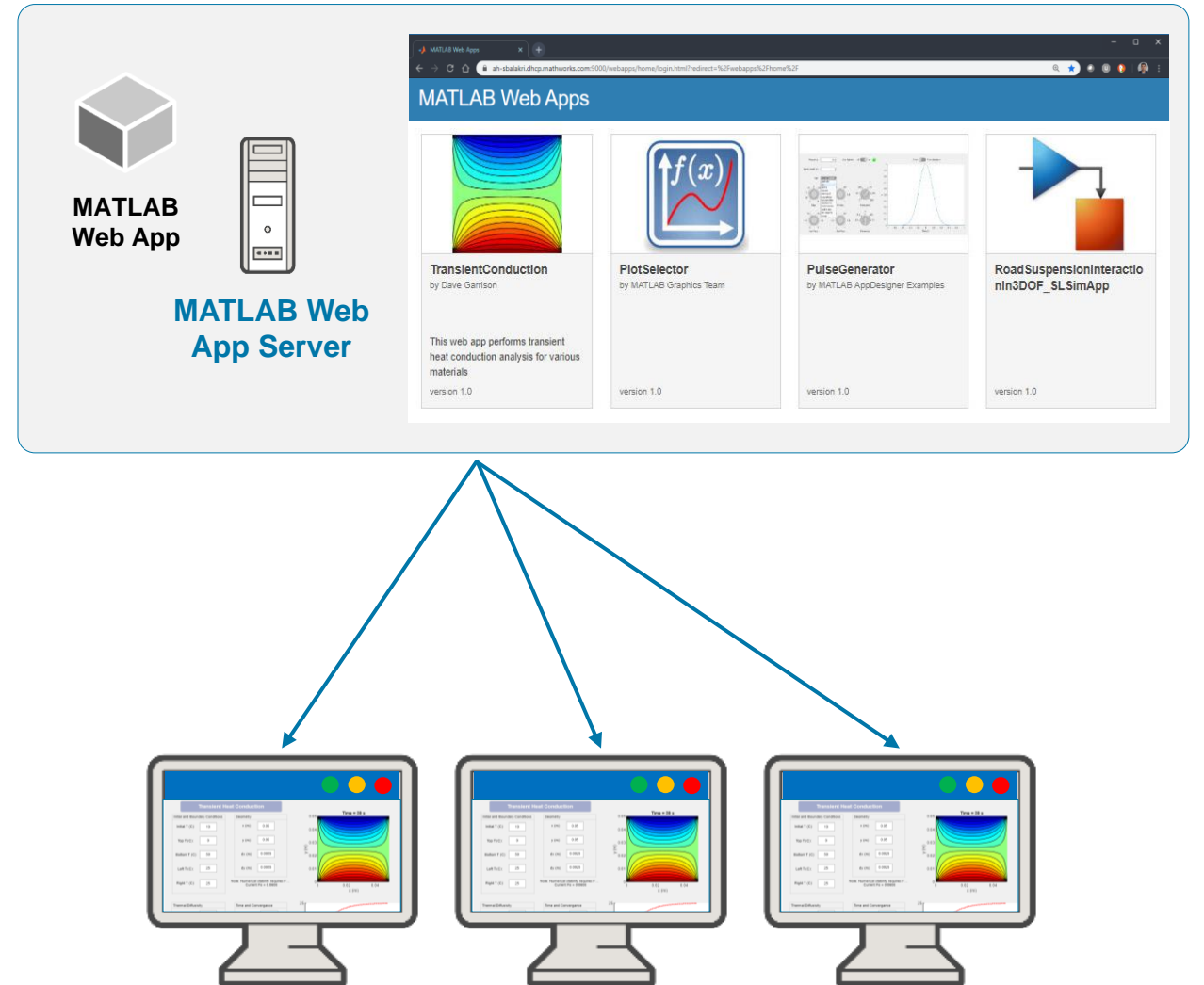
系统集成

- 与第三方语言开发的系统集成



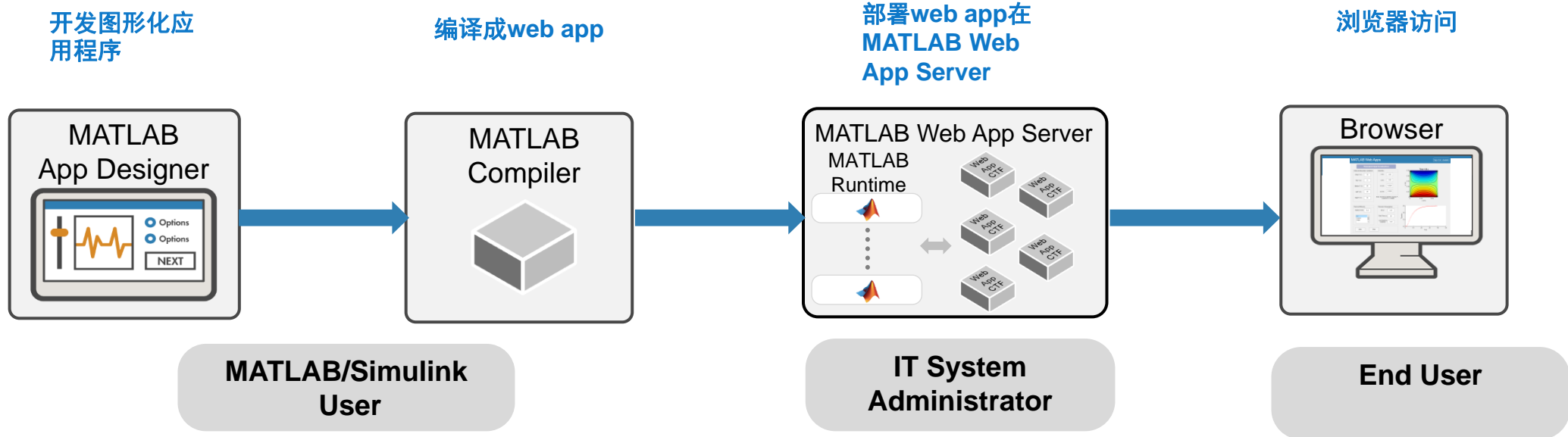
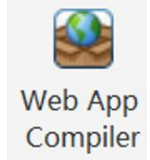
“云” 部署

- MATLAB Web App Server
在服务器端托管MATLAB应用程序
- Web应用程序（Web App）
在浏览器中运行的MATLAB应用程序。



MATLAB Web App Server - Web App

- 更加容易的管理和分享应用程序

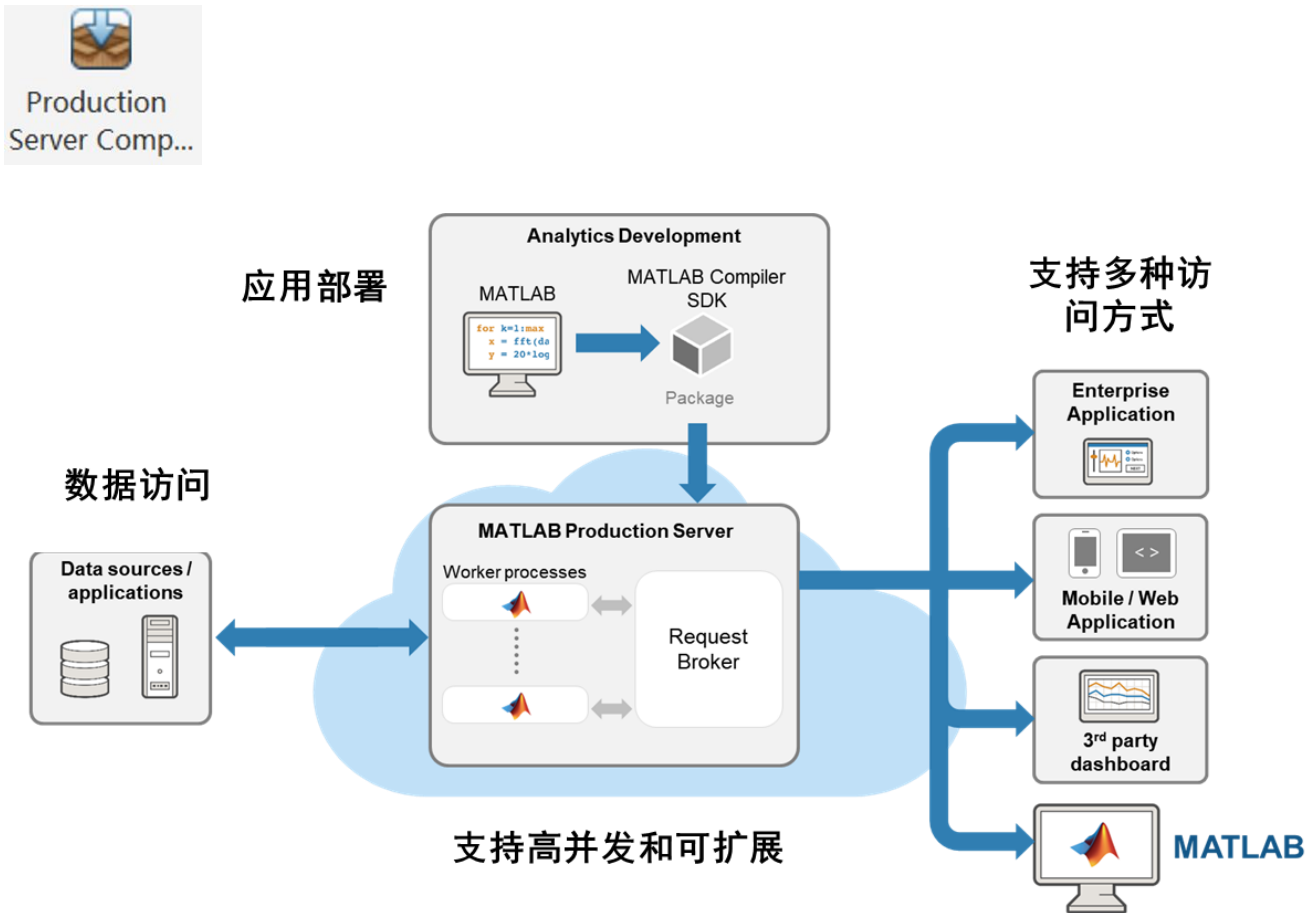


- 开发人员
- 共享应用程序

- IT人员
- 维护MATLAB Web App Server
- 管理Web App访问权限

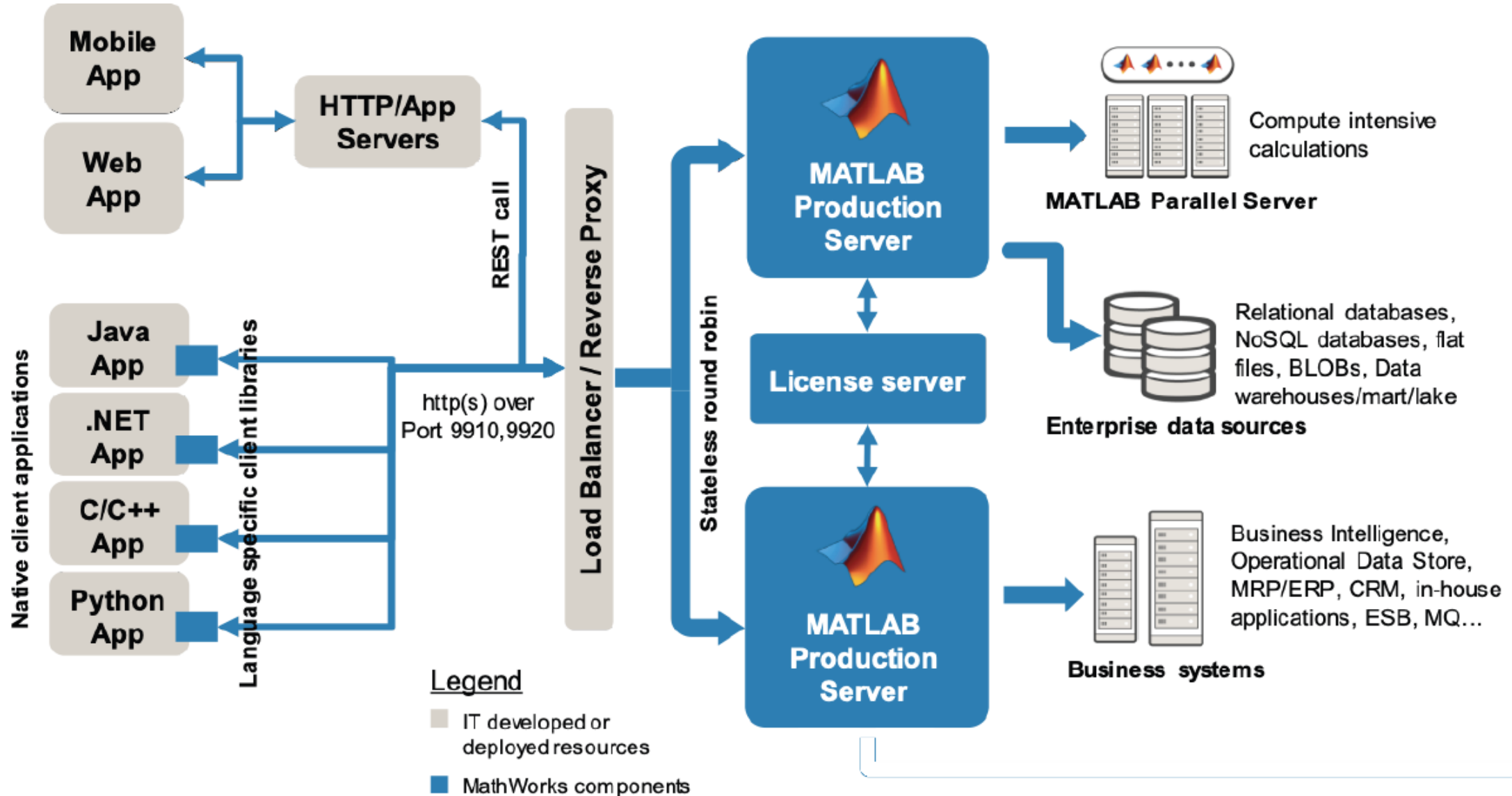
- 用户
- 避免本地安装应用程序和环境
- 通过浏览器随时随地可以访问应用程序

MATLAB Production Server (MPS) - Web API



- MPS是一个MATLAB代码远程运行环境。其可以通过图形化的方式安装在服务器上，并且独立运行。
- MPS自带图像化的管理界面，可以实现应用程序的配置、启动和停止，MATLAB代码的上传、部署和删除。
- MPS以Web API方式提供其上代码运行的调用接口，支持RESTful调用方式（JSON数据格式），支持JAVA，.NET，Python，C/C++等以客户端调用方式。
- MPS支持访问的高并发性，及服务器的可扩展性。

MATLAB Production Server支持的访问形式



MATLAB Production Server 应用场景

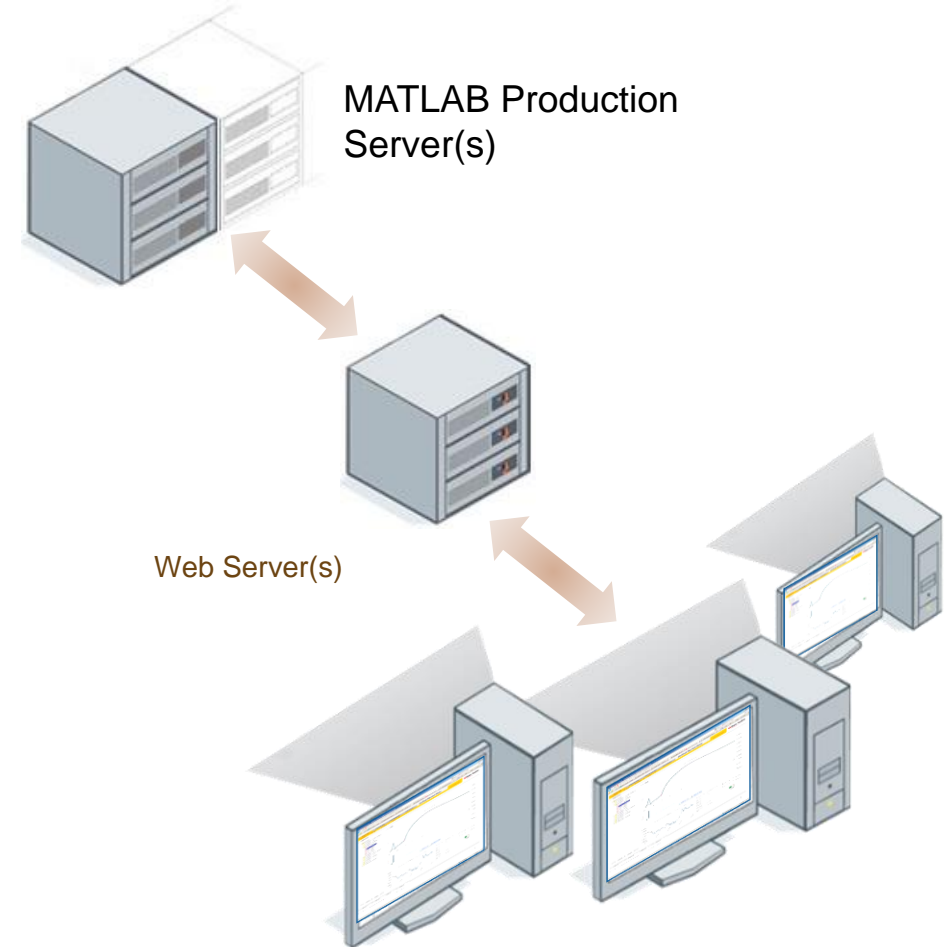
■ 服务后台

➤ 企业Web服务的后台

将MATLAB代码与企业Web服务器集成，
作为Web应用的计算引擎

➤ 物联网应用的服务后台

将MATLAB代码作为物联网的后台应用，实
现对传感器数据的处理。



Docker 微服务 (Microservice Docker Containers)



1

>> `compiler.build.producti`

Compile your code into a CTF (encrypted archive)

Deploy Industrial Cooling Fan Anomaly Detection Algorithm as Microservice

Create a microservice Docker® image from a predictive maintenance cooling fan algorithm.

[Open Live Script](#)



2

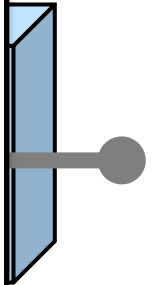
`image.microserv`

Docker container Windows, Runtime (for size) in the CT server

Deploy Wafer Map Defect Classifier as Microservice Using Docker

Create a microservice Docker® image from a wafer map defect classifier.

[Open Live Script](#)



HTTP RESTful Endpoint

push to send to cloud or rate repository

`docker run --rm -p 9900:9910 cartpoledqn-microservice -l trace &`

总结

- MATLAB Production Server
- MATLAB Web App Server
- MATLAB 可执行程序
- MATLAB 代码生成

2024 MathWorks 中国汽车年会

Thank you

