Deploying Cloud-Native Algorithms in Kubernetes

Nick Bonfatti, MathWorks

Pallavi Kar, MathWorks
Share the EXPO experience
#MATLABEXPO
How do you deploy your MATLAB algorithms today?

- Standalone executables
- Compiled shareable library
- Web Service
- Embedded Code
- Other
Optimizing cost with Simulation and Digital Twins

Carl Wouters

Link to User Story
Frontier Advisors Develops Web-Based Platform for Portfolio Analytics

“MATLAB and MATLAB Compiler SDK enabled us to rapidly deliver a sophisticated portfolio analytics web application with confidence that it will return accurate results extremely quickly, ensuring a highly usable and stable platform for our clients.”

– Lee Eriera, Frontier Advisors

Challenge

Provide clients with an industry-first web platform for portfolio modelling and analytics

Solution

Use MATLAB to develop and test analytics modules, and use MATLAB Compiler SDK to deploy them into a production .NET environment

Results

• Quantitative development decoupled from interface development
• Stable, responsive system deployed
• Rapid delivery of new features enabled

Learn about Frontier Advisors Technology

Link to User Story
Electricity Load and Price Forecasting Webinar Case Study

version 1.7.0.1 (12.3 MB) by Ameya Deoras

Slides and MATLAB® code for the day-ahead system load and price forecasting case study.

Overview  Functions  Examples  Reviews (25)  Discussions (79)

function y = loadForecast(date, temperature, isHoliday)

% LOADFORECAST performs a day-ahead load forecast using a pre-trained
% Neural-Network or Bagged Regression Tree model
%
% USAGE:
% y = loadForecast(model, date, hour, temperature, isWorkingDay))

% Process inputs
date = datenum(date);
if date < 7e5 % Convert from Excel numeric date to MATLAB numeric date if necessary
date = x2mdate(date);
end
Load forecasting deployed on MATLAB Production Server

Spotfire extension for MATLAB Production Server
Concurrent requests made to MATLAB Production Server
Understanding resource utilization using production server dashboard

Size your MATLAB Production Server
What to consider?

**Accessibility**
- Data
- Models
- APIs
- CI/CD

**Infrastructure**
- Compute
- Memory
- Security
- Network Access

**Maintenance & Recovery**
- Health Check
- Backup
- Server management

**Scalability**

**Vertical**
Bigger/Smaller server
High up-front cost vs risk of running out of resources

**Horizontal**
Number of nodes required concurrently
Load balancing across nodes/VMs
Regional and Global LBs

*Either way requires high maintenance if you keep on-premise!*
Key Takeaways

▪ Deploying MATLAB algorithms into cloud-native webservice using MATLAB Production Server

▪ Comparing VM based and Container based provisioning of MATLAB Production Server
  – Selecting deployment strategy based on requirements
  – Available reference architectures

▪ New Kubernetes-hosted MATLAB Production Server is
  – Performant
  – Resilient
  – Provides on-demand scaling
Which cloud platforms do you use?

- AWS
- Azure
- GCP
- Others
VM based reference architecture for MATLAB Production Server
Deploying to cloud using MATLAB Production Server

VM based MATLAB Production Server:
- Every instance is a separate VM
- Windows or Linux OS
- Manual scaling capability
- Web dashboard configuration

Reference Architectures available on GitHub:
- AWS
- Azure
- GCP
Key Takeaways

✓ Deploying MATLAB algorithms using MATLAB Production Server

✓ VM based provisioning of MATLAB Production Server on cloud

➢ Container based provisioning of MATLAB and Simulink models on the cloud

▪ New Kubernetes-hosted MATLAB Production Server is
  – Performant
  – Resilient
  – Provides on-demand scaling
Deploy MATLAB and Simulink algorithms in containers

Turn proof of concepts…

Into production web services deployed with DevOps principles

Package MATLAB models and Simulink simulations into a Docker container with RESTful HTTP endpoint(s) using the new microservice feature in MATLAB Compiler SDK R2022a

Create Microservice Docker Containers with MATLAB
Deploying to cloud using MATLAB Production Server

Container-based MATLAB Production Server
- Any Kubernetes cluster (Vendor independent)
- Lightweight, lower upfront infrastructure cost (New pods can be started quickly)
- Linux only
- Autoscaling
- CLI configuration

Reference Architectures available on GitHub:
- Any Kubernetes cluster, including AWS, Azure, and GCP
How can we manage containers?
How can we manage containers? Enter: Kubernetes

Kubernetes autoscaling based on requests

This makes it easy on solutions architect to design, plan and scale with a lot of flexibility
Deploying K8s hosted MATLAB Production Server
Which cloud architecture should we use?

We would like to standardize on containers where possible to avoid cloud vendor lock-in and have repeatable, automated deployments.

MATLAB Production Server has a Kubernetes-based deployment Reference Architecture to make that easy.

CTO
Drives operational strategy

System Architect
Deploys and operationalizes models on Azure cloud
Which cloud architecture should we use?

We have integrations that require Windows-based software

Process Engineer
Develops models in MATLAB and Simulink

In that case, we should use a VM-based MATLAB Production Server deployment

System Architect
Deploys and operationalizes models on Azure cloud
Key Takeaways

✓ Deploying MATLAB algorithms into cloud-native webservice using MATLAB Production Server

✓ Comparing VM based and Container based provisioning of MATLAB Production Server
  – Selecting deployment strategy based on requirements
  – Available reference architectures

✓ New Kubernetes-hosted MATLAB Production Server is
  – Performant
  – Resilient
  – Provides on-demand scaling
Call to Action

Explore the MATLAB Production Server Reference Architecture for Kubernetes

https://github.com/mathworks-ref-arch/matlab-production-server-on-kubernetes

• Uses existing MATLAB license server
• Works with cloud-managed or on-premise Kubernetes clusters
Thank you
Share the EXPO experience
#MATLABEXPO