

# Cloud / HPC for UM-Bridge Models

---

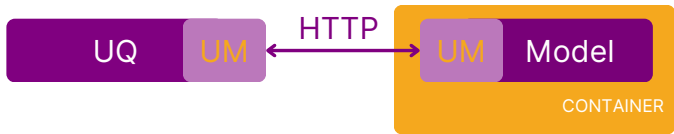
Linus Seelinger (Karlsruhe Institute of Technology, Germany)

Goal: "Automagically" scale up UM-Bridge models for challenging UQ problems; keep UQ client simple

Wish list:

- Avoid machine specific model setup
- Many instances of (in turn parallel) UM-Bridge models
- Single point of entry for UQ, load balancing

## UM-Bridge: Containerization - Portable Models



- Run tsunami model as easy as  
`docker run -p 4242:4242 linusseelinger/model-exahype-tsunami`
- Evaluate model in Python:  
`model = umbridge.HTTPModel('localhost:4242', 'forward')`  
`model([[0.1,0.4]])`

→ Separation of concerns!

Goal: "Automagically" scale up UM-Bridge models for challenging UQ problems; keep UQ client simple

Wish list:

- Avoid machine specific model setup → Containers! ✓
- Many instances of (in turn parallel) UM-Bridge models
- Single point of entry for UQ, load balancing

# HPC and Cloud Convergence

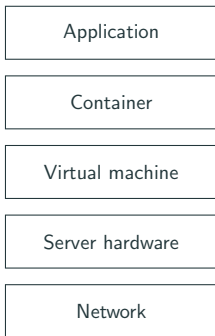
- HPC systems moving towards containers for portability
- Cloud systems moving towards HPC performance

⇒ HPC and cloud are (somewhat) converging

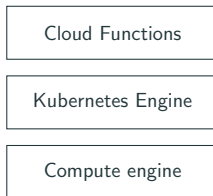
We go with cloud systems for excellent container support

# Overview: Cloud Infrastructure

## Infrastructure hierarchy



## Google Cloud Platform



Servers for rent, (very) different levels of abstraction possible

Kubernetes: "Container orchestration" - fully reproducible HPC setups

Goal: "Automagically" scale up UM-Bridge models for challenging UQ problems; keep UQ client simple

Wish list:

- Avoid machine specific model setup → Containers! ✓
- Many instances of (in turn parallel) UM-Bridge models → Cloud! ✓
- Single point of entry for UQ, load balancing



"Kubernetes, also known as K8s, is an open-source system for automating deployment, scaling, and management of containerized applications."

K8s setup defined through config files

→ Reusable configuration of UM-Bridge models, load balancers etc.

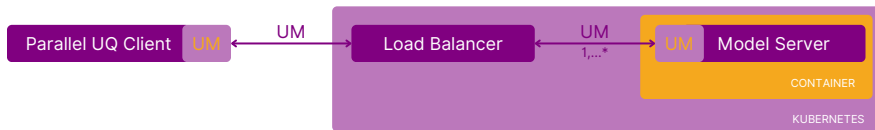


UM-Bridge currently provides two kubernetes configurations:

Parallel model instances,

- single node models
- multinode MPI parallel models

# Kubernetes Configuration - Sequential Model



Pre-built configuration, simply plug in your own model container

UQ client only sees an UM-Bridge server. But may make parallel requests!

Goal: "Automagically" scale up UM-Bridge models for challenging UQ problems; keep UQ client simple

Wish list:

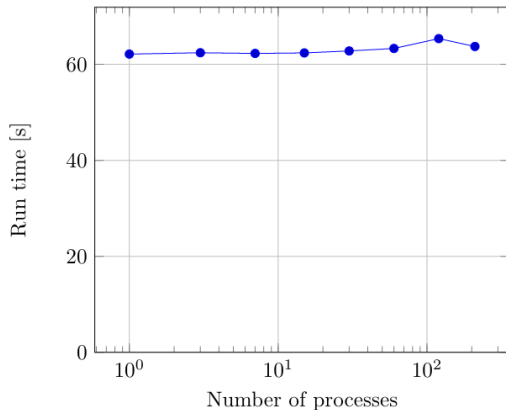
- Avoid machine specific model setup → Containers! ✓
- Many instances of (in turn parallel) UM-Bridge models → Cloud! ✓
- Single point of entry for UQ, load balancing  
→ UM-Bridge kubernetes config! ✓

## Model Configuration: model.yaml

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: model-deployment
spec:
  selector:
    matchLabels:
      app: model
  replicas: 4
  template:
    metadata:
      labels:
        app: model
    spec:
      containers:
      - name: model
        image: linusseelinger/model-l2-sea
        resources:
          requests:
            cpu: 1
            memory: 1Gi
          limits:
            cpu: 1
            memory: 1Gi
```

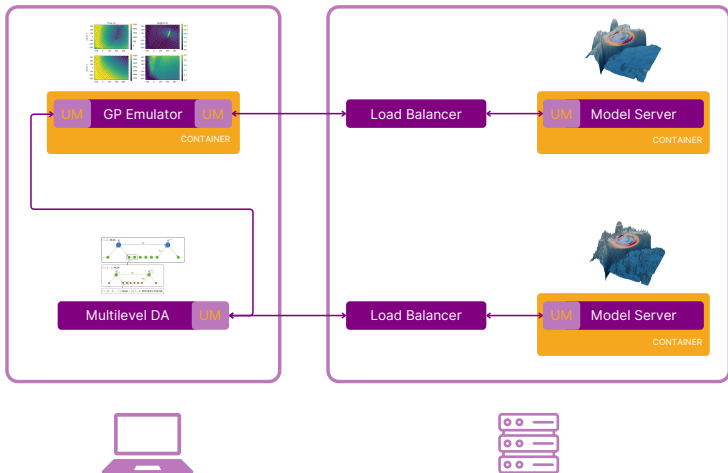
- **image:** Docker image with UM-Bridge model server (e.g. from Docker Hub)
- **replicas:** Number of model instances
- **requests / limits:** CPU and memory resources

# Scalability



- Model: L2-Sea
- Client: QMCPy with (unmodified!) thread parallelism
- Good weak scalability to 210 instances

# Example application



Successfully run on 2800 core Google Kubernetes Engine cluster

Demo: UM-Bridge Models on Google Kubernetes Engine

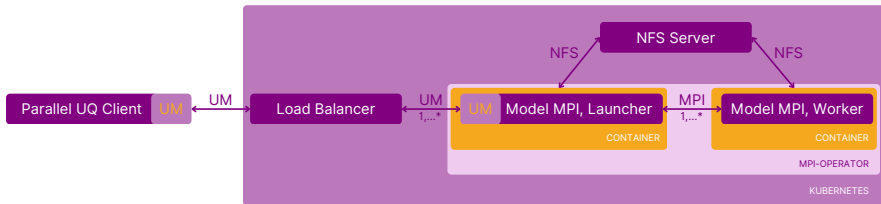
# MPI Parallel Models

---



- MPI *within* containers trivial, but restricted to one hardware node
- MPI *across* containers/nodes: Separate kubernetes configuration  
Requires `mpioperator` base image for model container

# Kubernetes Setup - MPI Parallel Model



Pre-built reference setup

Shared filesystem between nodes via NFS

Support for OpenMPI, Intel MPI, possibly MPICH in the future

Minor restrictions on model container

## Conclusions

---

- No model setup (due to containers)
- Reusable kubernetes configurations provided by UM-Bridge
- Upcoming: SLURM based HPC in addition to Cloud
- Simple UQ client can control HPC scale model