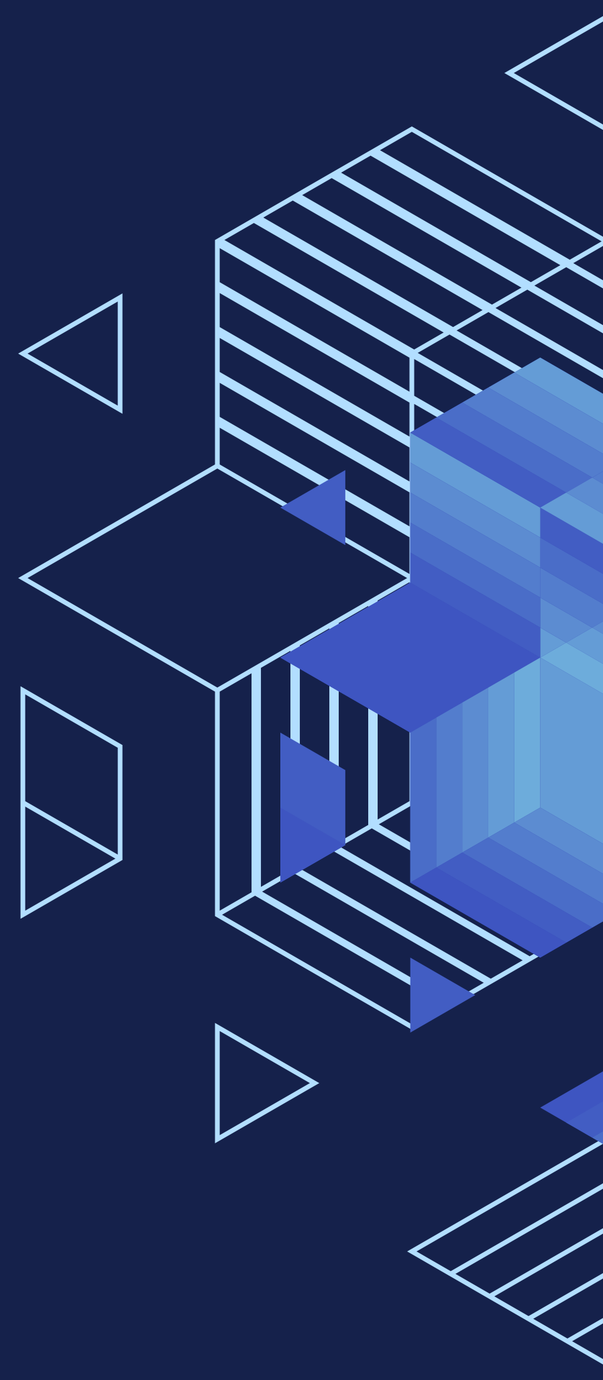
PROFILING HPC APPLICATIONS IN CONTAINERIZED ENVIRONMENTS

MASTER'S DEGREE IN COMPUTER ENGINEERING
HIGH PERFORMANCE COMPUTING AREA (M1.214)

ALBERT SANUY LOSTES
SERGIO ISERTE AGUT
JOSEP JORBA ESTEVE

Índex



1. Introducció

Motivació del projecte

Els computadors d'altes prestacions són essencials en l'actualitat donat que permeten processar informació i dur a terme operacions computacionals realment complexes a una velocitat molt alta.

Un del usos més habituals és en l'estudi de paradigmes de caràcter científic. No obstant això, en les últimes dècades, algunes investigacions han fracassat a causa de la dificultat per a poder reproduir-les amb un alt nivell de fiabilitat.

1. Introducció

Objectius

L'objectiu d'aquest estudi és construir una imatge emprant Docker, que contindrà una implementació de l'estàndard MPI, una aplicació distribuïda d'àmbit científic i una eina que permetrà generar les traces de l'execució per a una posterior anàlisi.

Aquestes traces permetran comparar el rendiment per a diferents càrregues de treball que es duran a terme.

2. Arquitectura

Dispositius

Els dispositius emprats en aquest estudi són dues Raspberries Pi 4 Model B, cada una d'elles connectada a una font d'alimentació, i a un switch, model TP-LINK TL-SG1005D, que permet la comunicació entre aquestes.

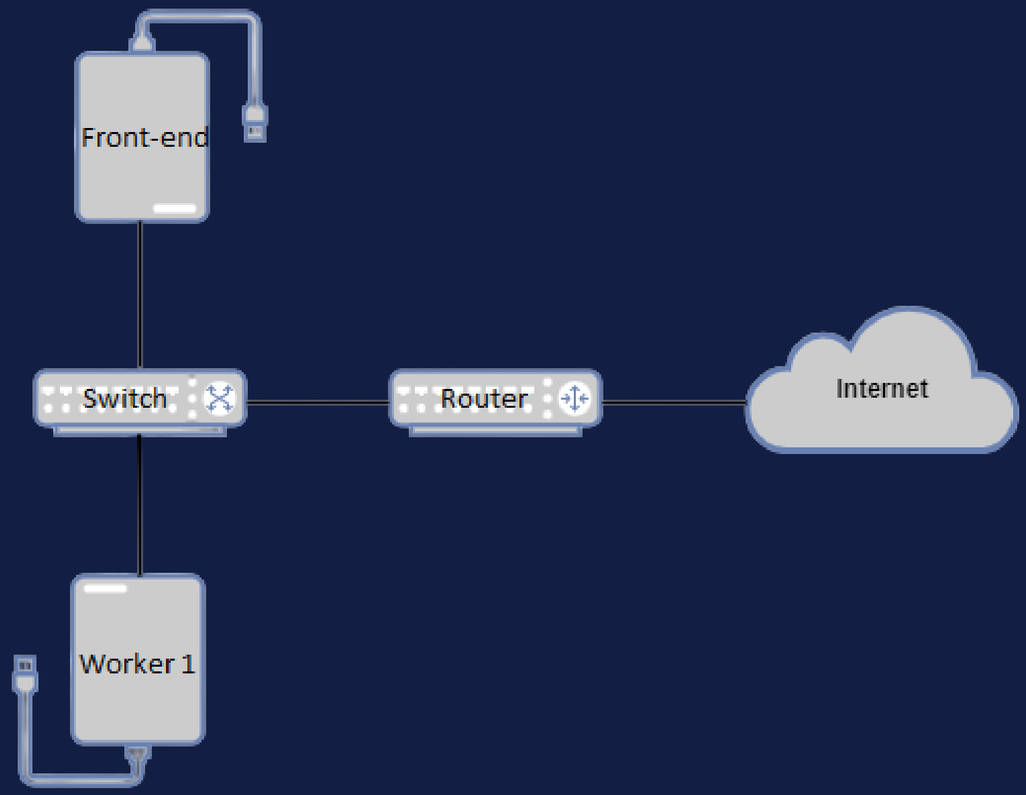
Specifications	
Processor	Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
RAM	2GB LPDDR4-3200 SDRAM
Network	2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0, BLE Gigabit Ethernet
Hard drive	16GB micro-SD card for loading operating system and data storage
Screen ports	2 × micro-HDMI ports (up to 4kp60 supported)

2. Arquitectura

Disseny del sistema

El switch i el router actuen en diferents capes del model OSI.

Mentre que el switch treballa en la capa d'enllaç de dades i connecta els dispositius, el router treballa en la capa de xarxa i s'encarrega de l'encaminament entre una o més xarxes.



3. Eines

OpenMPI

Implementació de l'estàndard MPI.

OpenFOAM

Aplicació distribuïda d'àmbit científic.

Extrae

Eina encarregada de generar les traces.

Paraver

Interfície emprada per a visualitzar les traces.

4. Contenidors

Docker

Servei que permet crear aplicacions web en paquets anomenats contenidors.

Singularity

Servei que permet la virtualització a nivell de sistema operatiu i enfocat a clústers HPC.

5. Creació de la imatge

1. Compilació de les llibreries
2. Crear imatge de Docker
3. Publicar imatge de Docker
4. Singularity Image File
5. Execució

6. Execució

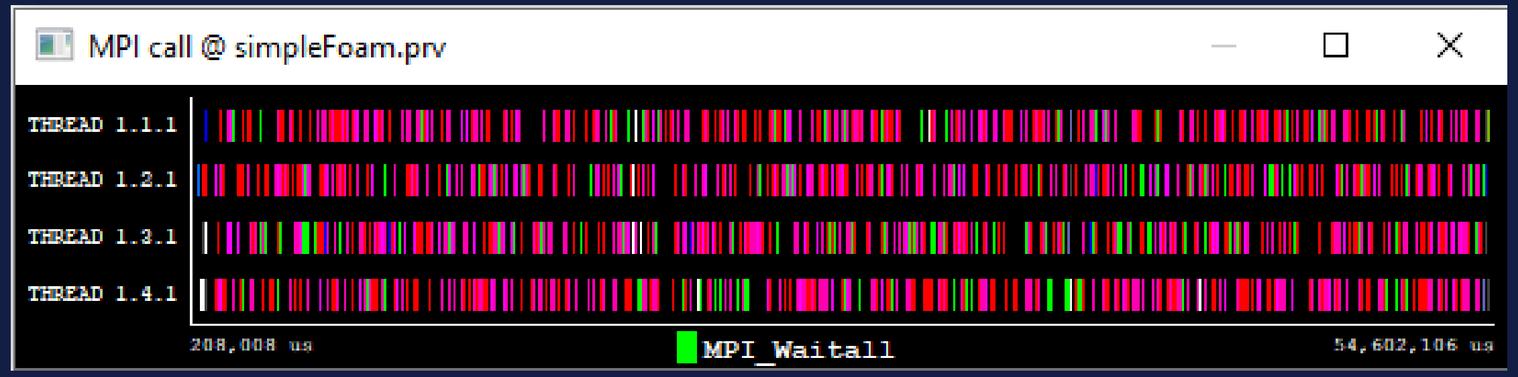
La simulació d'OpenFOAM escollida per a l'estudi és un cas dels exemples que permet comparar diferents models de turbulències.

En primer lloc, s'executarà la comanda que permet generar la geometria del domini en un conjunt de tres dimensions. A continuació, s'executarà la comanda que permet descompondre el domini en tasques més petites que puguin ser executades paral·lelament. Finalment, s'iniciarà la simulació en mode distribuït.

```
> mpirun -n 4 -host node0:2,node1:2 -quiet --mca btl_tcp_if_include 192.168.0.1/24,192.168.0.2/24
  singularity exec --writable-tmpfs /SHARED/openmpi.img ./run.sh
```

7. Anàlisi

Finalment, com a resultat de l'execució de l'aplicació, i després que Extrae finalitzi el procés d'unir les traces intermèdies, es generaran tres fitxers: la traça, un fitxer de configuració i un fitxer que conté la distribució de l'aplicació segons l'ús dels recursos del clúster.



MPI call profile @ simpleFoam.prv #3

Default

	Outside MPI	MPI_Send	MPI_Recv	MPI_Isend	MPI_Irecv	MPI_Waitall	MPI_Allreduce	MPI_Alltoall	MPI_Comm_rank	MPI_Comm_size	MPI_Comm_free	MPI_Init	MPI_Finalize	MPI_Probe
THREAD 1.1.1	91.75 %	0.03 %	0.10 %	2.23 %	1.12 %	1.67 %	3.04 %	0.05 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.01 %
THREAD 1.2.1	87.43 %	0.02 %	0.27 %	1.62 %	0.88 %	6.19 %	3.35 %	0.10 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.13 %
THREAD 1.3.1	87.45 %	0.03 %	0.31 %	1.74 %	1.00 %	5.88 %	3.35 %	0.11 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.13 %
THREAD 1.4.1	91.79 %	0.02 %	0.16 %	2.40 %	1.17 %	1.84 %	2.43 %	0.06 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.13 %
Total	358.42 %	0.10 %	0.85 %	7.98 %	4.17 %	15.58 %	12.18 %	0.32 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.40 %
Average	89.61 %	0.02 %	0.21 %	2.00 %	1.04 %	3.90 %	3.05 %	0.08 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.10 %
Maximum	91.79 %	0.03 %	0.31 %	2.40 %	1.17 %	6.19 %	3.35 %	0.11 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.13 %
Minimum	87.43 %	0.02 %	0.10 %	1.62 %	0.88 %	1.67 %	2.43 %	0.05 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.01 %
StDev	2.16 %	0.01 %	0.08 %	0.32 %	0.11 %	2.14 %	0.38 %	0.02 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.05 %
Avg/Max	0.98	0.75	0.68	0.83	0.89	0.63	0.91	0.75	0.75	0.86	0.93	0.77	0.78	0.74

8. Conclusions

- Core principles: Reproducibility and reliability
- Benefits of using containers
- MPI programming paradigm as a solution to solve compelx problems
- Instrumentation tools
- Future improvements

GRÀCIES