Enhancing Parallel Computing CPU and I/O Performance through Malleable Resource Management

Paula Sanchez-Checa, Genaro Sanchez-Gallegos, Javier García-Blas,
Jesús Carretero and David E. Singh

uc3m | Universidad Carlos III de Madrid

ecdc EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

isc Instituto de Salud Carlos III

RES RED ESPAÑOLA DE SUPERCOMPUTACIÓN

GOBIERNO DE ESPAÑA MINISTERIO DE SANIDAD, CONSUMO Y BIENESTAR SOCIAL

HPC-IODC Workshop: HPC I/O in the Data Center. ISC 2025

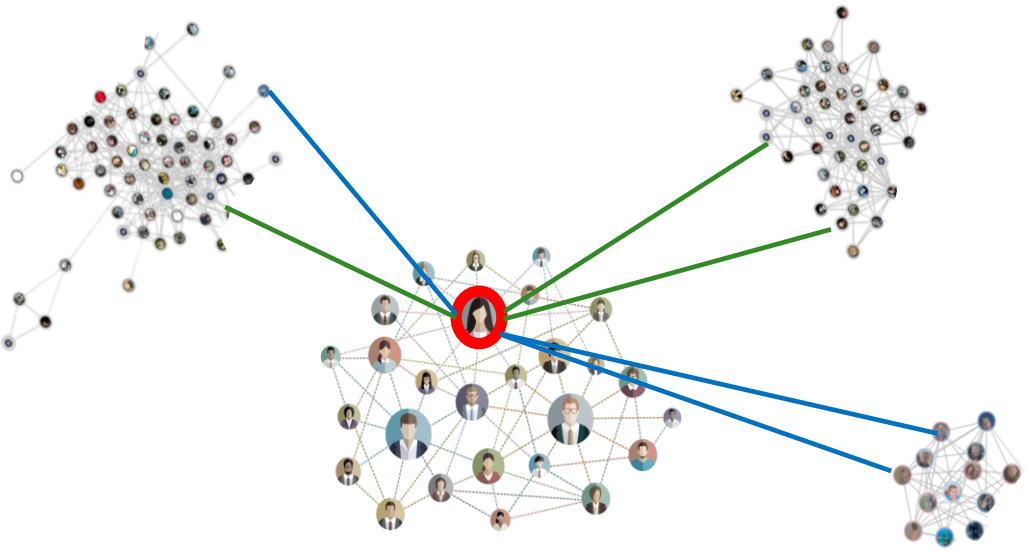
Background

- EpiGraph
  - Agent-based epidemiologic simulator
  - Multiple COVID-19 variants
  - Extended to model Influenza and RSV
  - Social model built from social networks and contact matrices
  - Connections reproduce social particularities: day of week, time of day, profession, touristic season
  - Vaccination, waning immunity, etc.

epigraph.uc3m.es

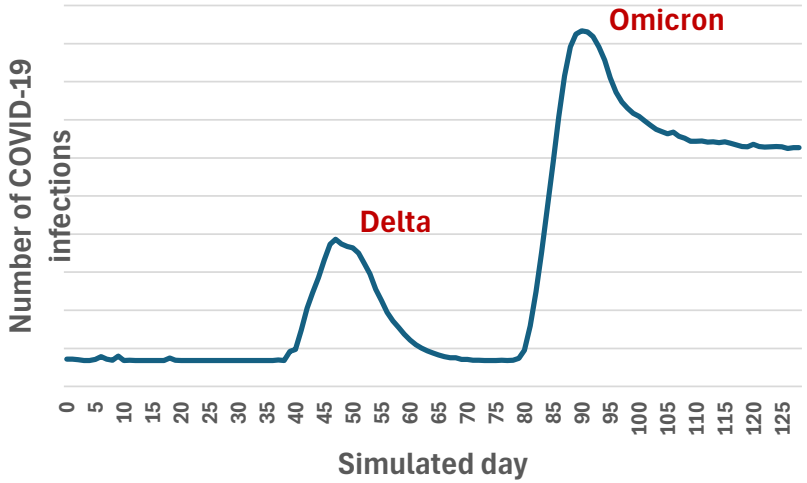
Background

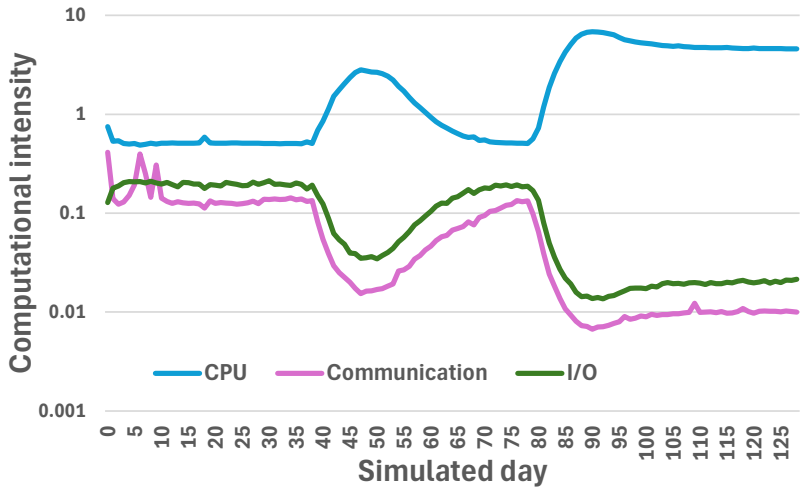
- EpiGraph simulation scenario
  - SARS-CoV-2 Omicron variant outbreak in Spain in November 2021
  - Madrid city
  - 3.5 million individuals



Background

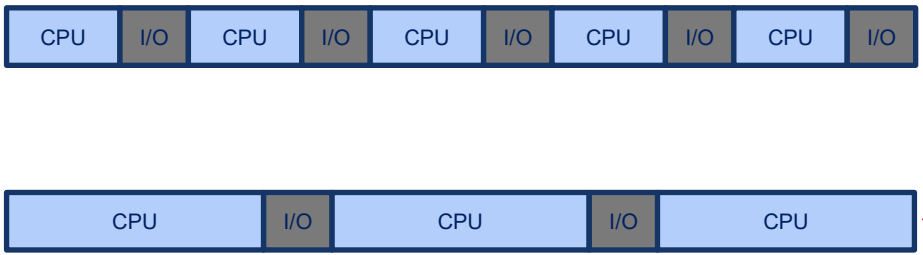
- EpiGraph alternates
  - CPU: 10 minutes (simulated time)
  - Communication: every simulated hour
  - I/O: every simulated day

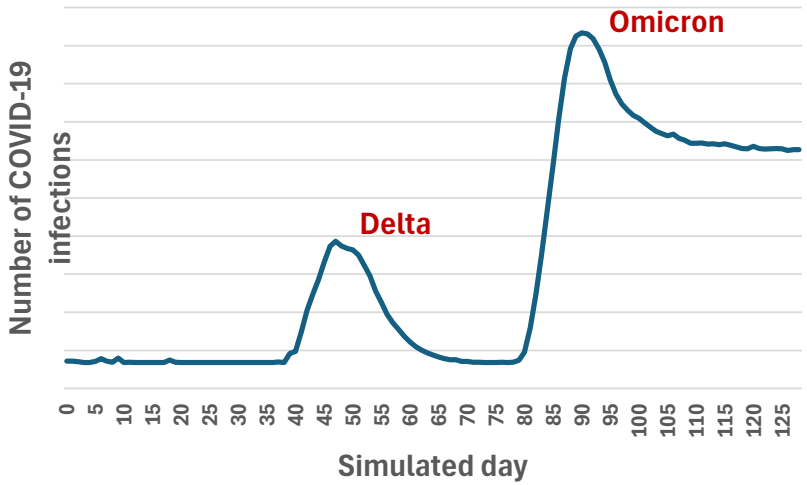


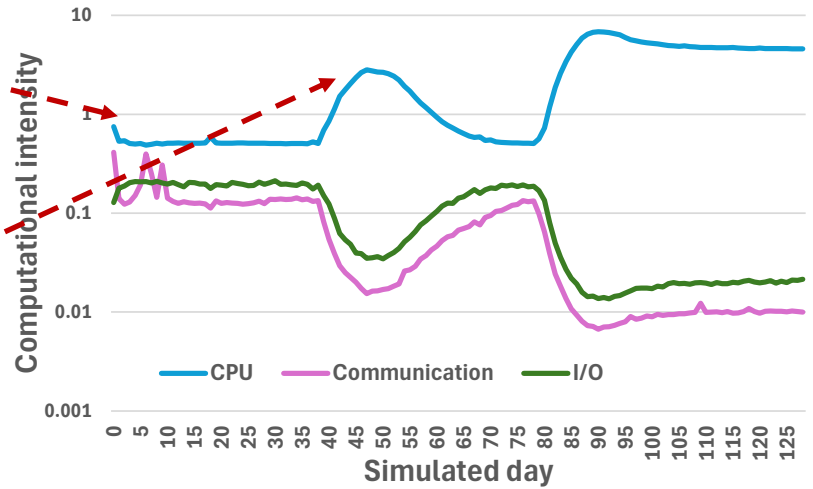


Background

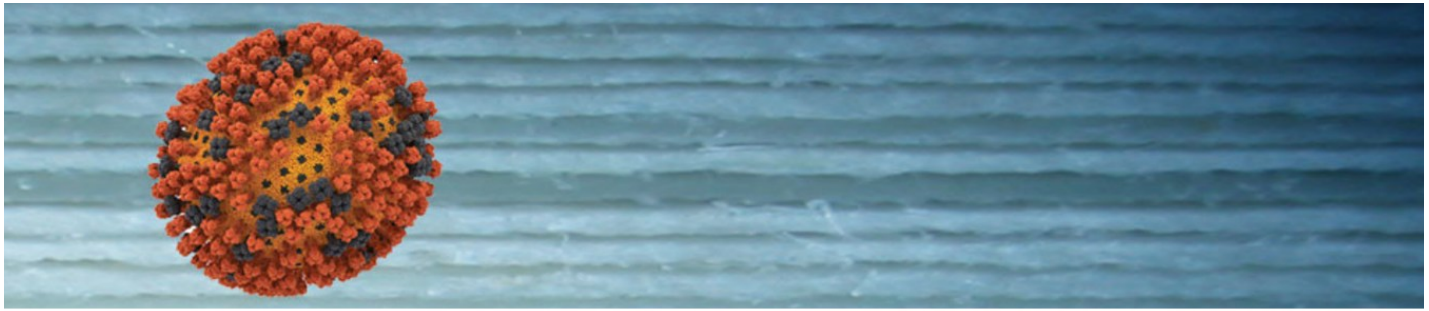
- EpiGraph alternates
  - CPU: 10 minutes (simulated time)
  - Communication: every simulated hour
  - I/O: every simulated day







Background



Combining Malleability and Distributed Control Mechanisms to Reduce I/O Contention

Paula Sanchez-Checa1, Javier Garcia-Blas, Jesus Carretero, and David E. Singh

ISC 2025 IXPUG Workshop

uc3m | Universidad Carlos III de Madrid

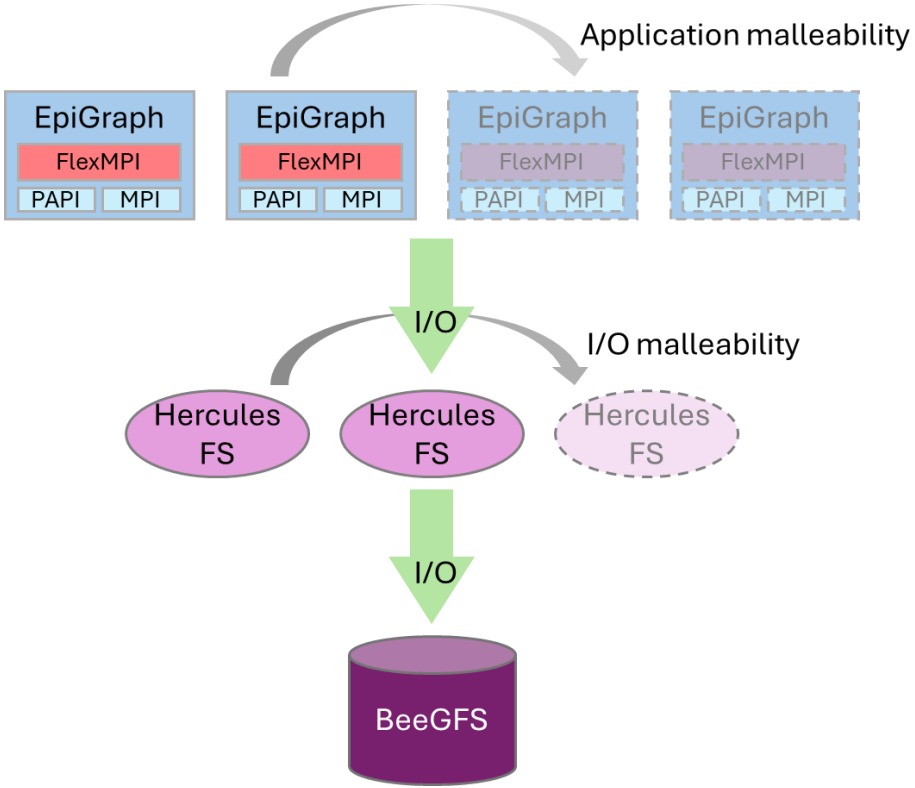
ecdc EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL

isc Instituto de Salud Carlos III

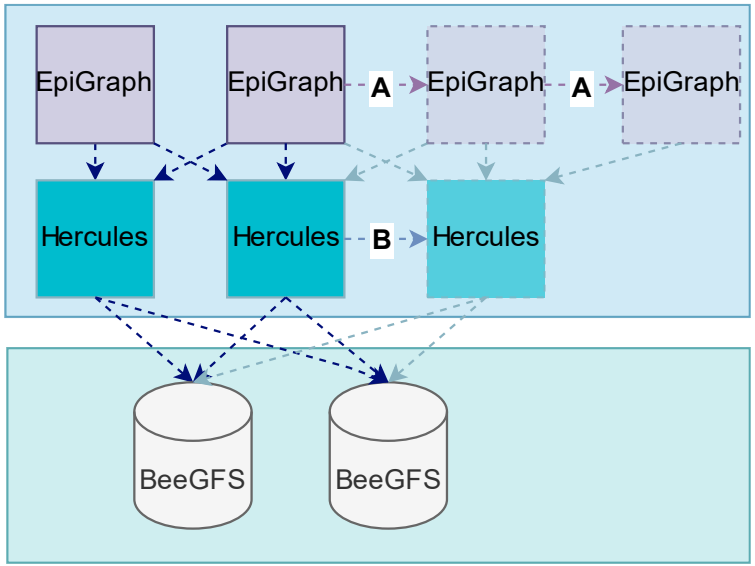
RES RED ESPAÑOLA DE SUPERCOMPUTACIÓN

GOBIERNO DE ESPAÑA | MINISTERIO DE SANIDAD, CONSUMO Y BIENESTAR SOCIAL

Background



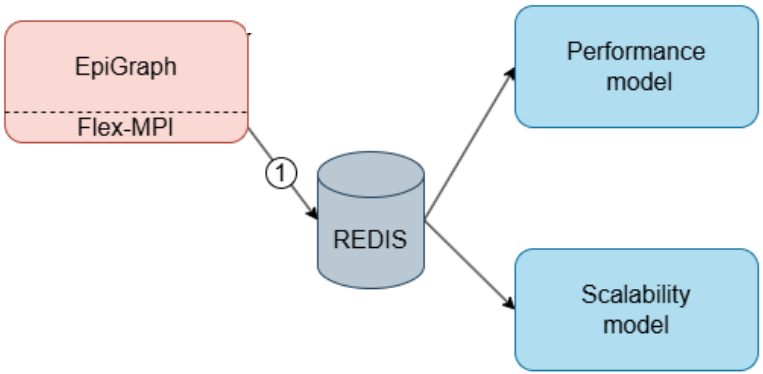
- Hercules
  - Ad-hoc in-memory filesystem
  - Dynamic capabilities



Architecture

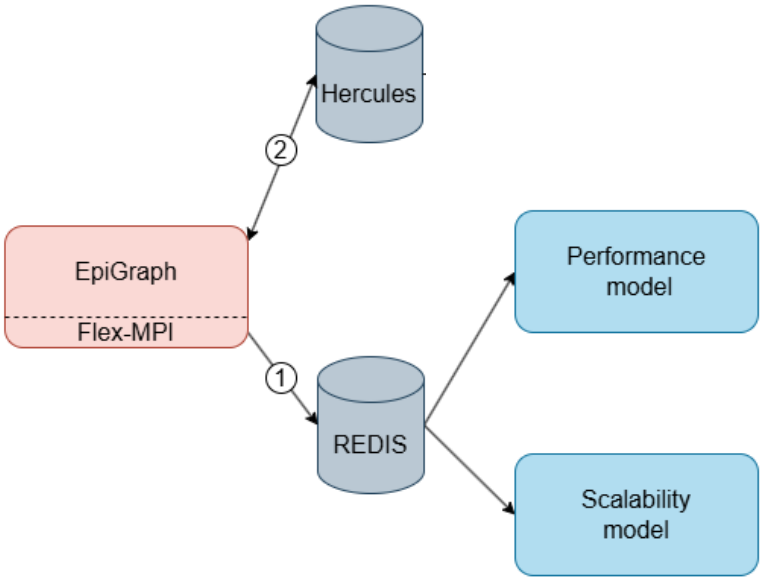
Model generation

1. Application monitoring and trace generation



Architecture

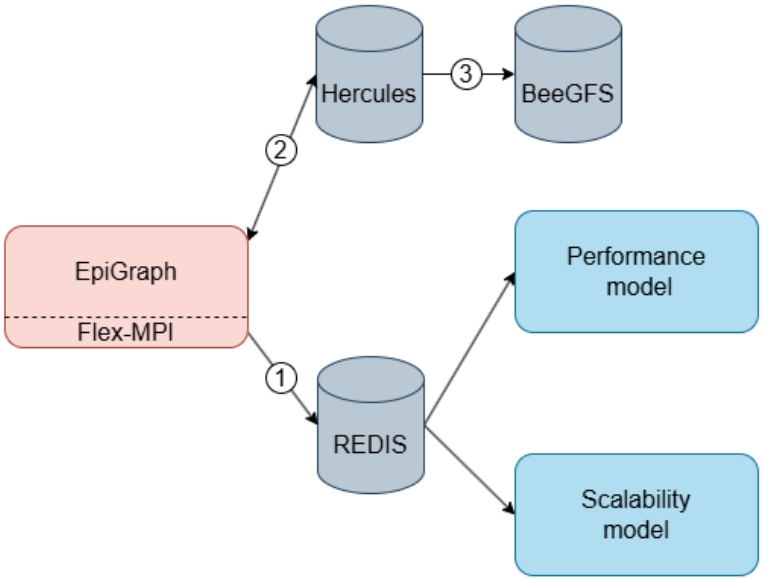
Model generation



1. Application monitoring and trace generation
2. Ad-hoc FS

Architecture

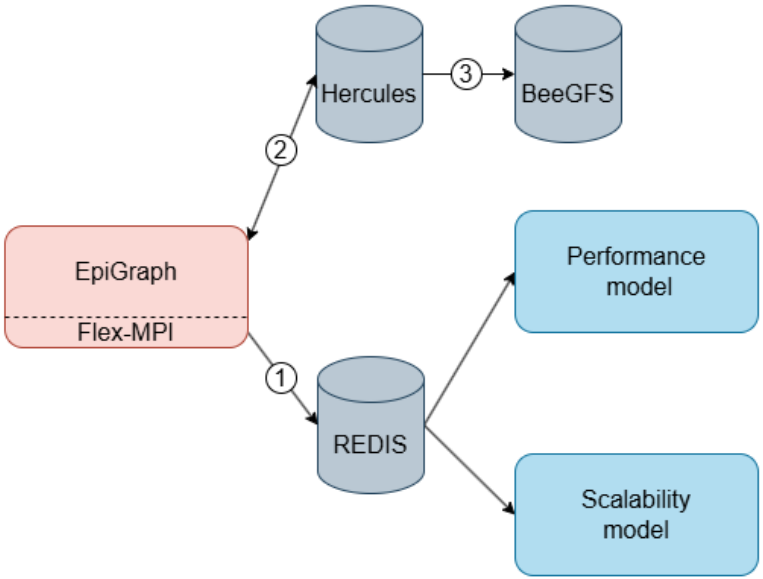
Model generation



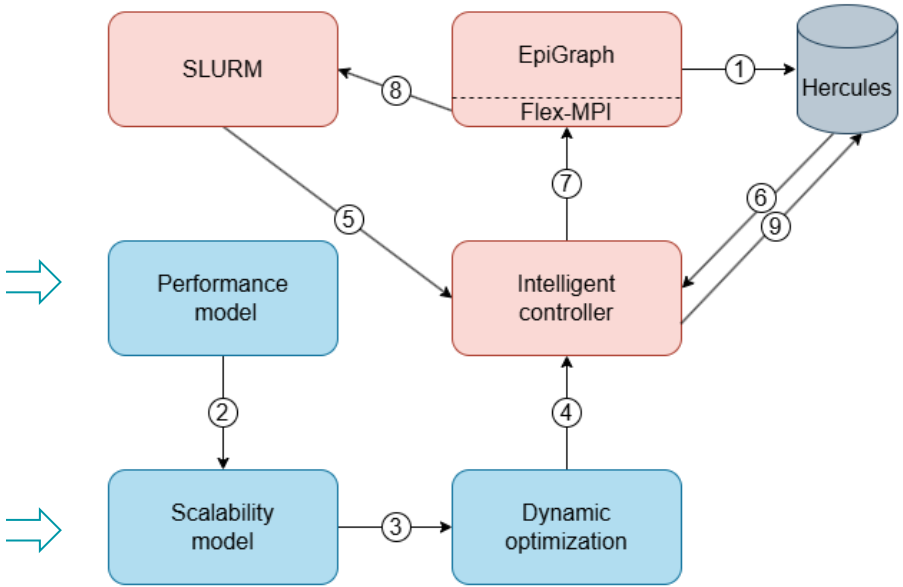
1. Application monitoring and trace generation
2. Ad-hoc FS
3. Back-end FS

Architecture

Model generation

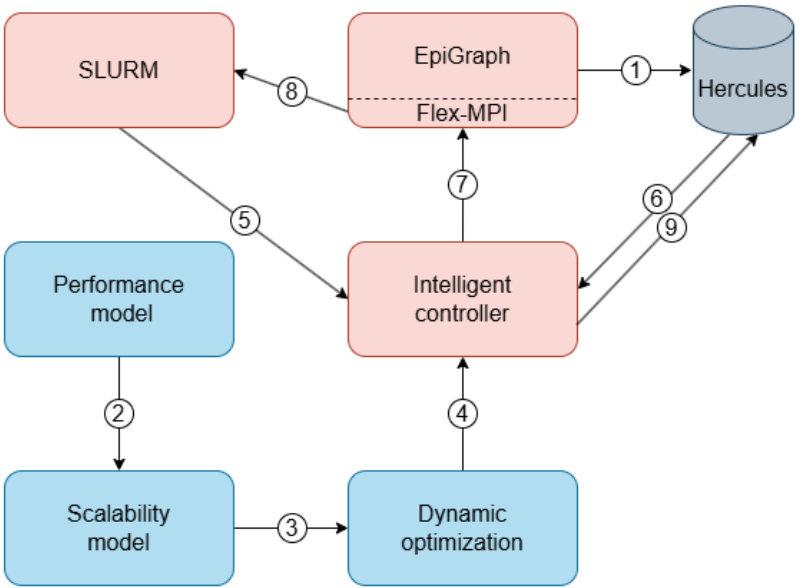


Dynamic optimization



Architecture

Dynamic optimization



Dynamic optimization

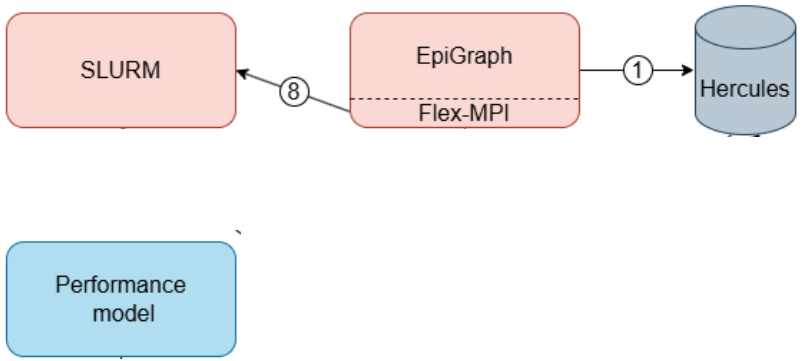
1. I/O access on Hercules

EpiGraph execution

Dynamic optimization

1. I/O access on Hercules
2. Performance prediction

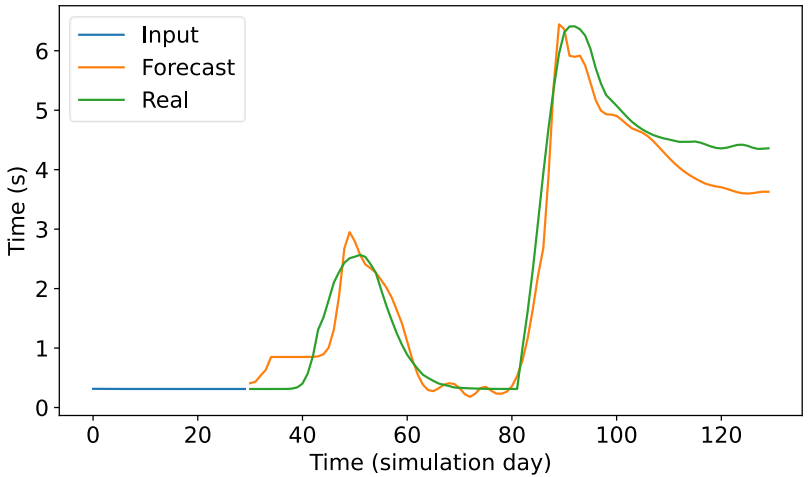
EpiGraph execution



Dynamic optimization

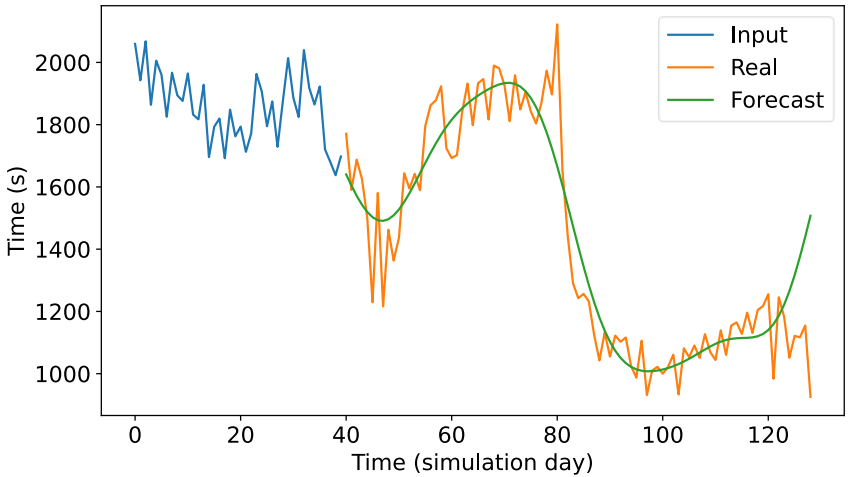
CPU time

- Long Short-Term Memory NN
- RMSE: 5,5



Communication time

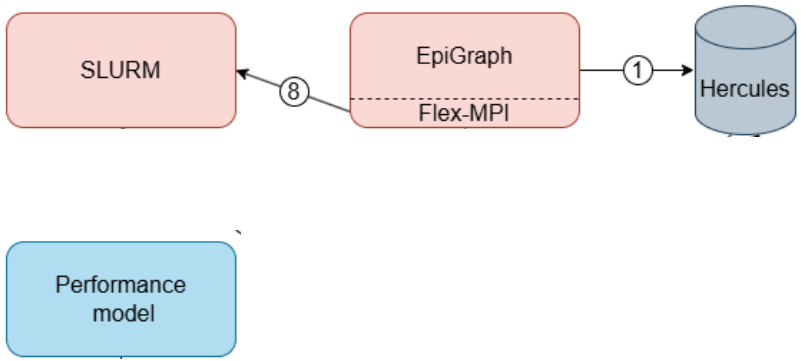
- Low-pass filter
- RMSE: 1,3



Dynamic optimization

1. I/O access on Hercules
2. Performance prediction

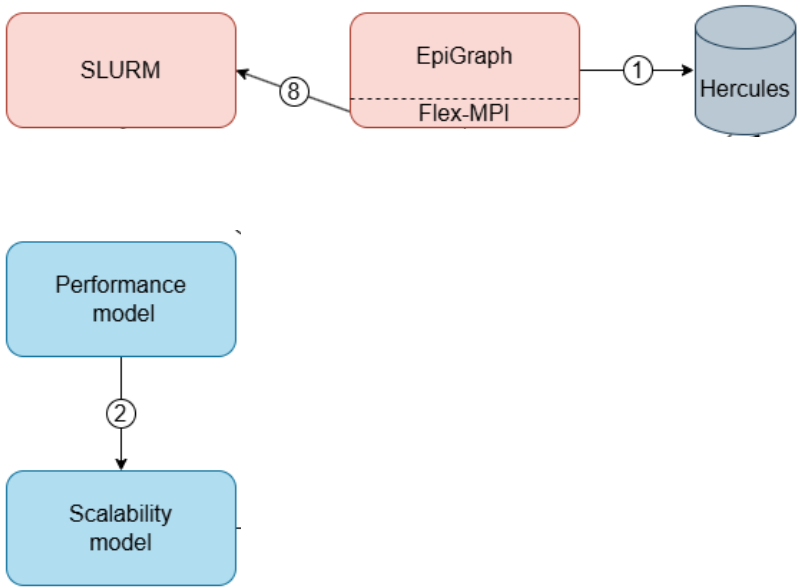
EpiGraph execution



Dynamic optimization

1. I/O access on Hercules
2. Performance prediction
3. Scalability model

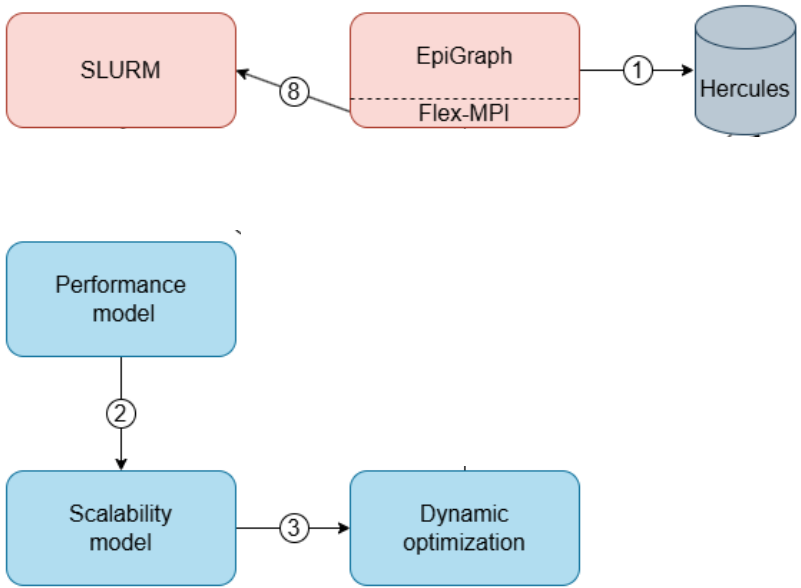
EpiGraph execution



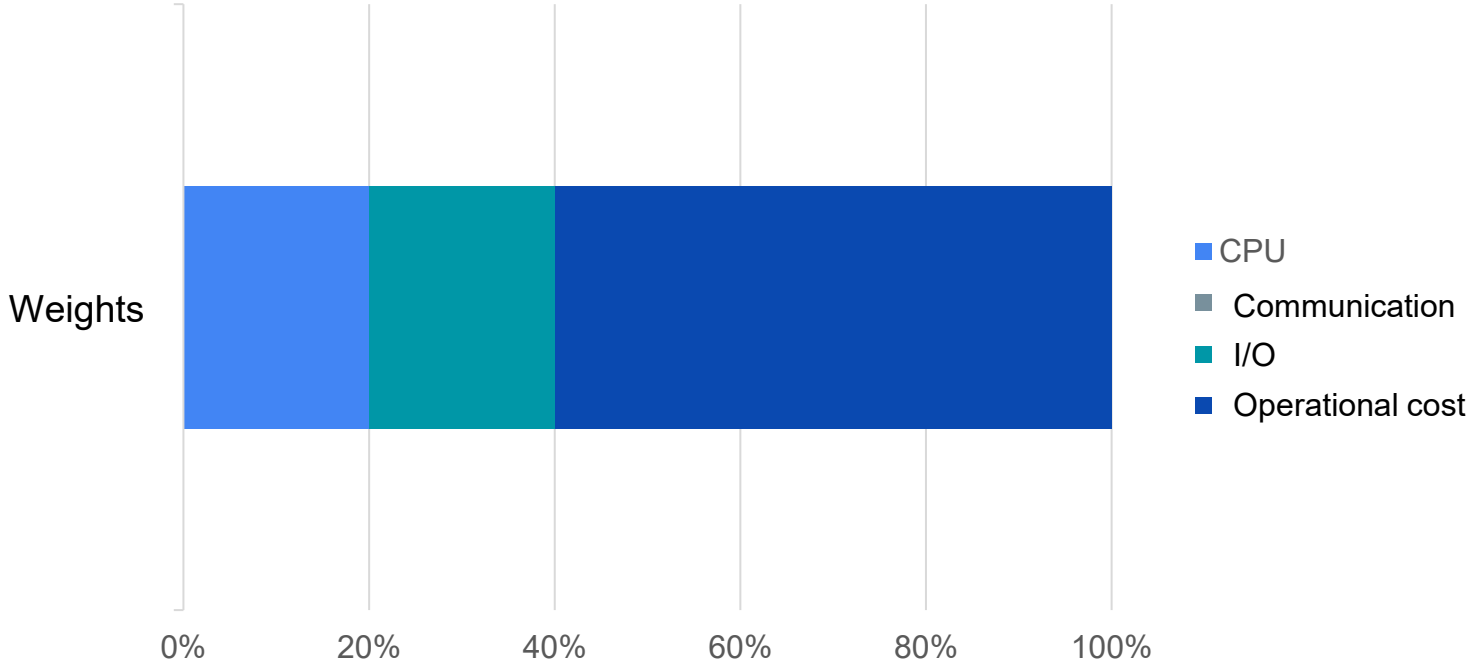
Dynamic optimization

1. I/O access on Hercules
2. Performance prediction
3. Scalability model
4. Configuration points

EpiGraph execution



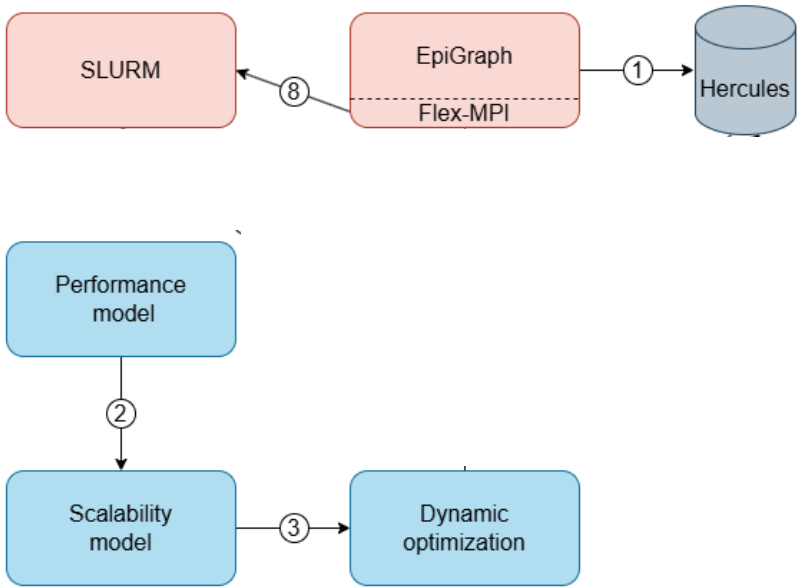
$$\begin{cases} c_t^i = \alpha \cdot t_{cpu} + \beta \cdot t_{com} + \gamma \cdot t_{e/s} + \delta \cdot t_{op} \\ \alpha + \beta + \gamma + \delta = 1 \end{cases}$$



Dynamic optimization

1. I/O access on Hercules
2. Performance prediction
3. Scalability model
4. Configuration points

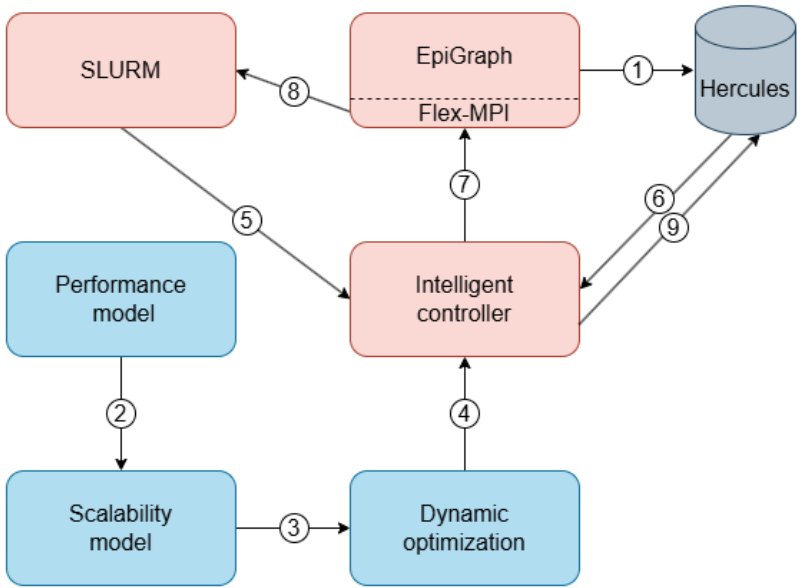
EpiGraph execution



Dynamic optimization

1. I/O access on Hercules
2. Performance prediction
3. Scalability model
4. Configuration points
5. Optimization process

EpiGraph execution



Evaluation

- HPC4AI Laboratory, Universidad di Torino

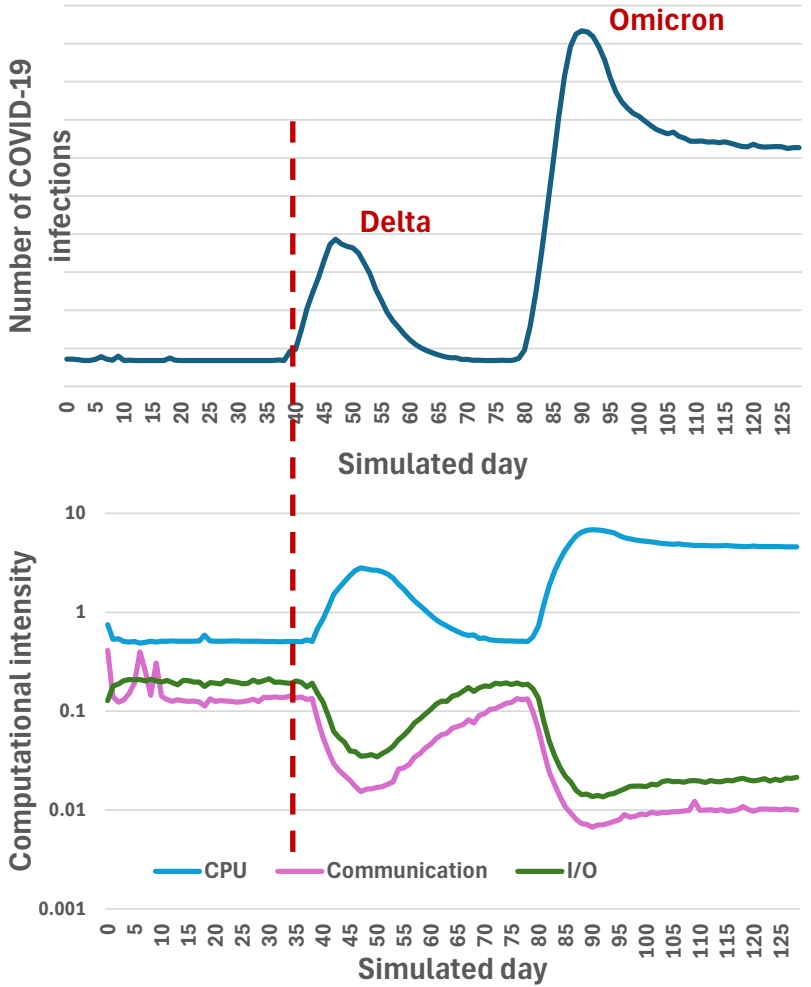
Id	Initial configuration	(α,	β,	γ,	δ)	Execution time (s)	Operational time (s)	Reconfigurations
1	**(1, 2)**	**(-,**	**-,**	**-,**	**-)**	**507**	**2.447**	———
2	(1, 2)	(1,	0,	0,	0)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
3	(1, 2)	(0,	0,	1,	0)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
4	(1, 2)	(0.4,	0,	0.4,	0.2)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
5	(1, 2)	(0.2,	0,	0.2,	0.6)	325	1.651	$(1, 2) \xrightarrow{40} (5, 4) \xrightarrow{84} (5, 2)$
6	(1, 2)	(0.1,	0,	0.1,	0.8)	294	1.547	$(1, 2) \xrightarrow{41} (5, 2)$
7	**(2, 4)**	**(-,**	**-,**	**-,**	**-)**	**483**	**2.835**	———
8	(2, 4)	(1,	0,	0,	0)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
9	(2, 4)	(0,	0,	1,	0)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
10	(2, 4)	(0.4,	0,	0.4,	0.2)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
11	(2, 4)	(0.2,	0,	0.2,	0.6)	321	1.747	$(2, 4) \xrightarrow{42} (5, 4) \xrightarrow{84} (5, 2)$
12	(2, 4)	(0.1,	0,	0.1,	0.8)	305	1.633	$(2, 4) \xrightarrow{42} (3, 4) \xrightarrow{84} (5, 2)$
13	**(8, 4)**	**(-,**	**-,**	**-,**	**-)**	**379**	**4.210**	———
14	(8, 4)	(1,	0,	0,	0)	379	4.210	(8, 4)
15	(8, 4)	(0,	0,	1,	0)	379	4.210	(8, 4)
16	(8, 4)	(0.4,	0,	0.4,	0.2)	379	4.210	(8, 4)
17	(8, 4)	(0.2,	0,	0.2,	0.6)	383	2.472	$(8, 4) \xrightarrow{40} (5, 4) \xrightarrow{84} (5, 2)$
18	(8, 4)	(0.1,	0,	0.1,	0.8)	382	2.495	$(8, 4) \xrightarrow{40} (3, 4) \xrightarrow{82} (5, 2)$

Evaluation

- HPC4AI Laboratory, Universidad di Torino

Id	Initial configuration	(α,	β,	γ,	δ)	Execution time (s)	Operational time (s)	Reconfigurations
1	**(1, 2)**	**(-,**	**-,**	**-,**	**-)**	**507**	**2.447**	
2	(1, 2)	(1,	0,	0,	0)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
3	(1, 2)	(0,	0,	1,	0)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
4	(1, 2)	(0.4,	0,	0.4,	0.2)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
5	(1, 2)	(0.2,	0,	0.2,	0.6)	325	1.651	$(1, 2) \xrightarrow{40} (5, 4) \xrightarrow{84} (5, 2)$
6	(1, 2)	(0.1,	0,	0.1,	0.8)	294	1.547	$(1, 2) \xrightarrow{41} (5, 2)$
7	**(2, 4)**	**(-,**	**-,**	**-,**	**-)**	**483**	**2.835**	
8	(2, 4)	(1,	0,	0,	0)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
9	(2, 4)	(0,	0,	1,	0)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
10	(2, 4)	(0.4,	0,	0.4,	0.2)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
11	(2, 4)	(0.2,	0,	0.2,	0.6)	321	1.747	$(2, 4) \xrightarrow{42} (5, 4) \xrightarrow{84} (5, 2)$
12	(2, 4)	(0.1,	0,	0.1,	0.8)	305	1.633	$(2, 4) \xrightarrow{42} (3, 4) \xrightarrow{84} (5, 2)$
13	**(8, 4)**	**(-,**	**-,**	**-,**	**-)**	**379**	**4.210**	
14	(8, 4)	(1,	0,	0,	0)	379	4.210	(8, 4)
15	(8, 4)	(0,	0,	1,	0)	379	4.210	(8, 4)
16	(8, 4)	(0.4,	0,	0.4,	0.2)	379	4.210	(8, 4)
17	(8, 4)	(0.2,	0,	0.2,	0.6)	383	2.472	$(8, 4) \xrightarrow{40} (5, 4) \xrightarrow{84} (5, 2)$
18	(8, 4)	(0.1,	0,	0.1,	0.8)	382	2.495	$(8, 4) \xrightarrow{40} (3, 4) \xrightarrow{82} (5, 2)$

Evaluation

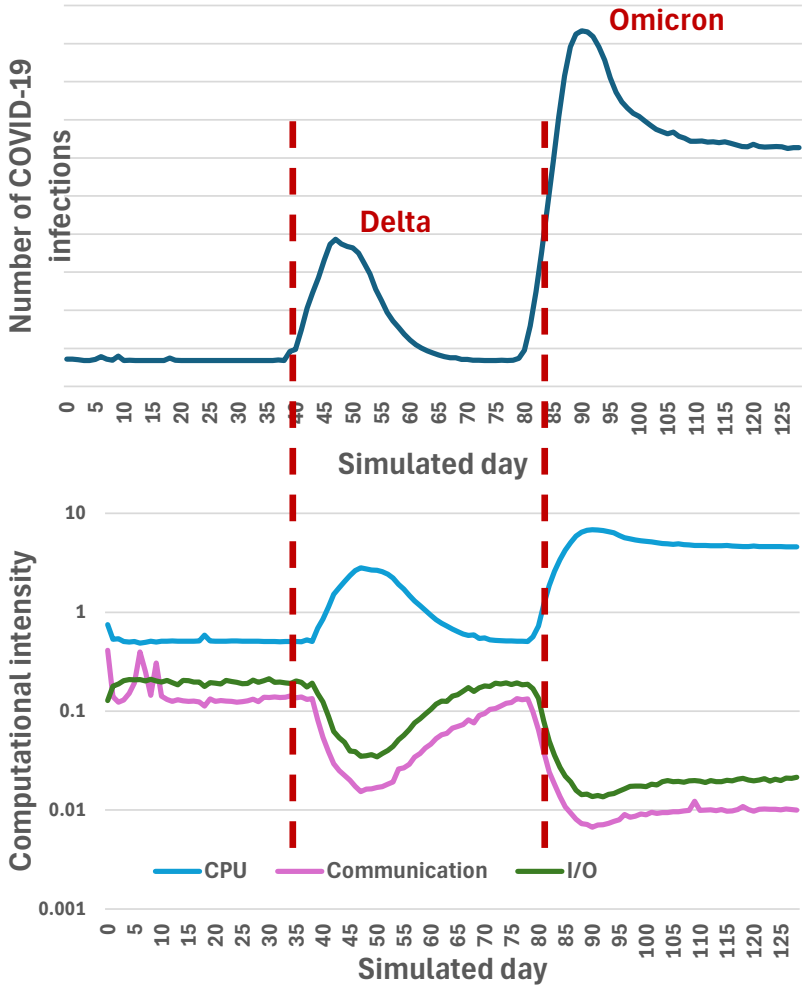

Omicron
Delta
Number of COVID-19 infections
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125
Simulated day
10
1
0.1
0.01
0.001
Computational intensity
CPU
Communication
I/O
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125
Simulated day

Evaluation

- HPC4AI Laboratory, Universidad di Torino

Id	Initial configuration	(α,	β,	γ,	δ)	Execution time (s)	Operational time (s)	Reconfigurations
1	**(1, 2)**	**(-,**	**-,**	**-,**	**-)**	**507**	**2.447**	——
2	(1, 2)	(1,	0,	0,	0)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
3	(1, 2)	(0,	0,	1,	0)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
4	(1, 2)	(0.4,	0,	0.4,	0.2)	338	3.000	$(1, 2) \xrightarrow{40} (8, 4)$
5	(1, 2)	(0.2,	0,	0.2,	0.6)	325	1.651	$(1, 2) \xrightarrow{40} (5, 4) \xrightarrow{84} (5, 2)$
6	(1, 2)	(0.1,	0,	0.1,	0.8)	294	1.547	$(1, 2) \xrightarrow{41} (5, 2)$
7	**(2, 4)**	**(-,**	**-,**	**-,**	**-)**	**483**	**2.835**	——
8	(2, 4)	(1,	0,	0,	0)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
9	(2, 4)	(0,	0,	1,	0)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
10	(2, 4)	(0.4,	0,	0.4,	0.2)	329	3.145	$(2, 4) \xrightarrow{40} (8, 4)$
11	(2, 4)	(0.2,	0,	0.2,	0.6)	321	1.747	$(2, 4) \xrightarrow{42} (5, 4) \xrightarrow{84} (5, 2)$
12	(2, 4)	(0.1,	0,	0.1,	0.8)	305	1.633	$(2, 4) \xrightarrow{42} (3, 4) \xrightarrow{84} (5, 2)$
13	**(8, 4)**	**(-,**	**-,**	**-,**	**-)**	**379**	**4.210**	——
14	(8, 4)	(1,	0,	0,	0)	379	4.210	(8, 4)
15	(8, 4)	(0,	0,	1,	0)	379	4.210	(8, 4)
16	(8, 4)	(0.4,	0,	0.4,	0.2)	379	4.210	(8, 4)
17	(8, 4)	(0.2,	0,	0.2,	0.6)	383	2.472	$(8, 4) \xrightarrow{40} (5, 4) \xrightarrow{84} (5, 2)$
18	(8, 4)	(0.1,	0,	0.1,	0.8)	382	2.495	$(8, 4) \xrightarrow{40} (3, 4) \xrightarrow{82} (5, 2)$

Evaluation

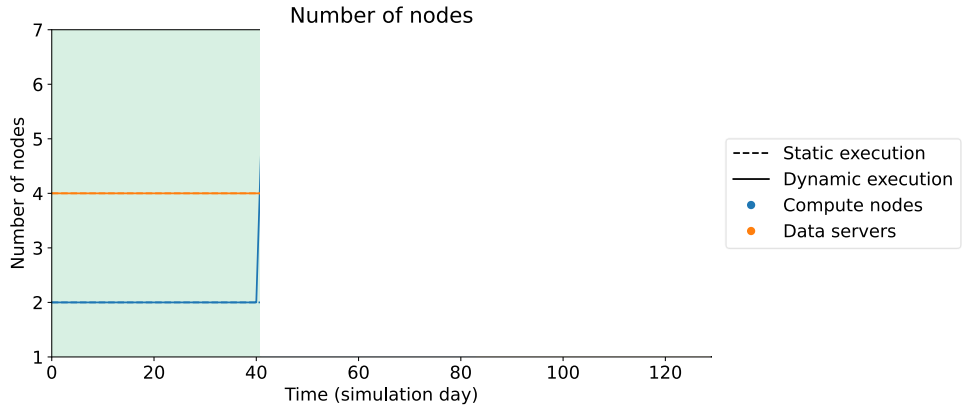

Omicron
Delta
Number of COVID-19 infections
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125
Simulated day
10
1
0.1
0.01
0.001
Computational intensity
CPU
Communication
I/O
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125
Simulated day

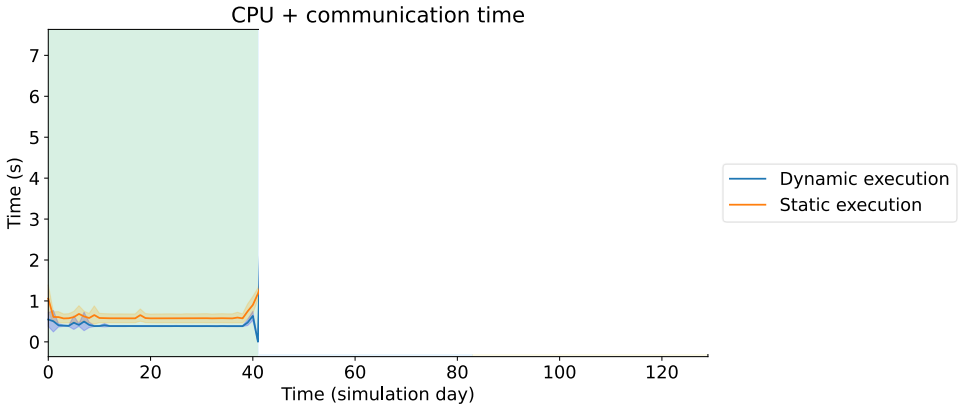
Evaluation

CPU vs communication times

$$(2,\ 4) \xrightarrow{42} (5,\ 4) \xrightarrow{84} (5,\ 2)$$

	Static execution	Dynamic execution
Execution time	321 s	483 s
Operating time	1.747 s	2.835 s



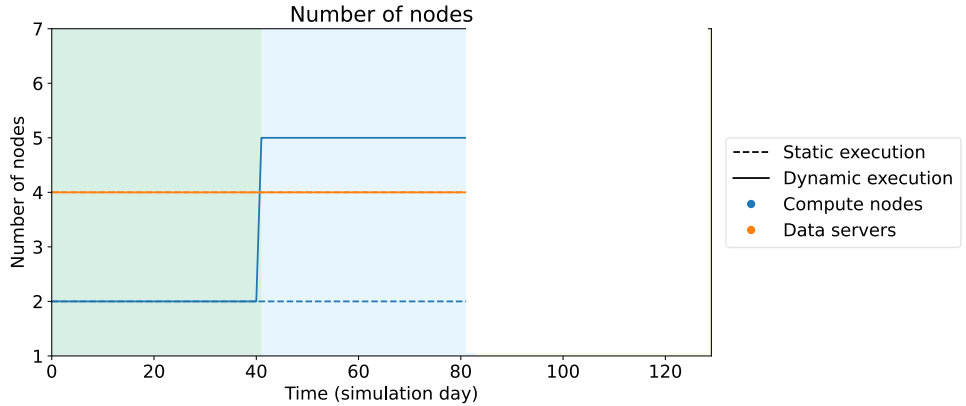


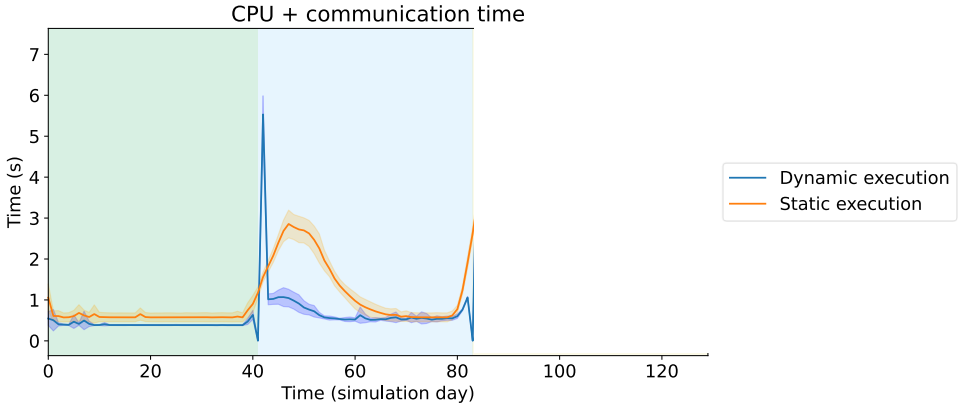
Evaluation

CPU vs communication times

$$(2,\ 4) \xrightarrow{42} (5,\ 4) \xrightarrow{84} (5,\ 2)$$

	Static execution	Dynamic execution
Execution time	321 s	483 s
Operating time	1.747 s	2.835 s



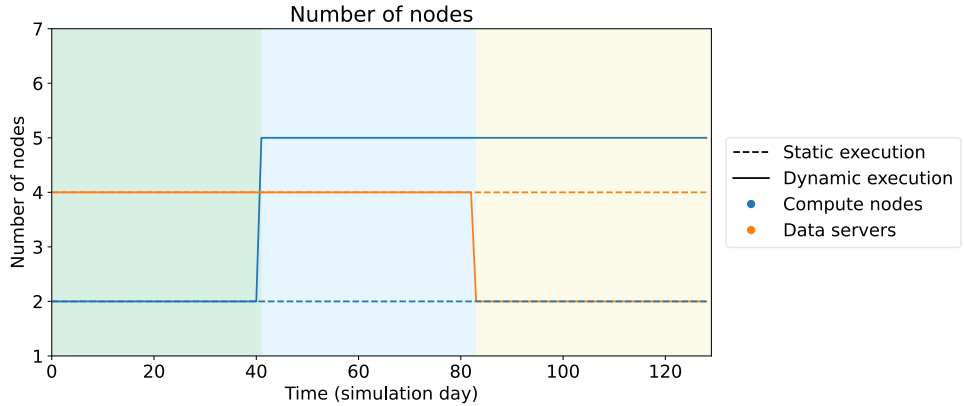


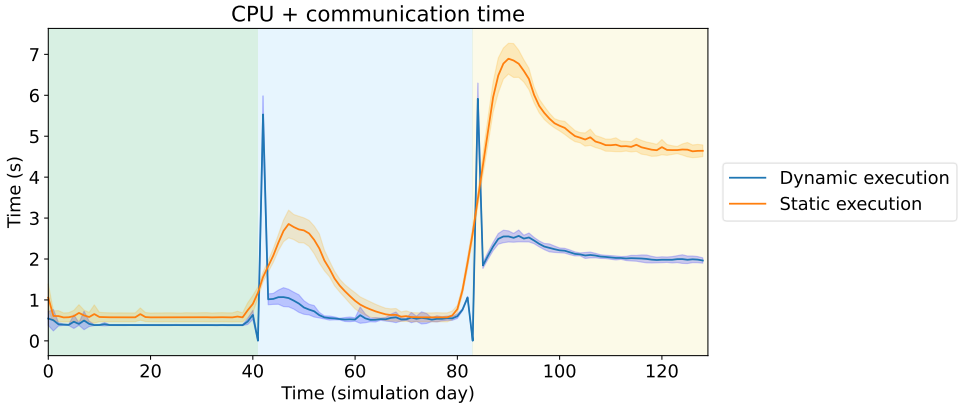
Evaluation

CPU vs communication times

$$(2,\ 4) \xrightarrow{42} (5,\ 4) \xrightarrow{84} (5,\ 2)$$

	Static execution	Dynamic execution
Execution time	321 s	483 s
Operating time	1.747 s	2.835 s



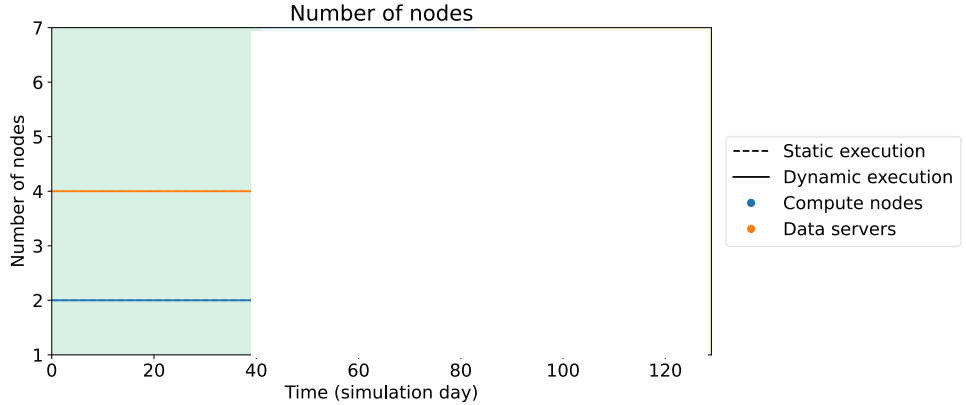


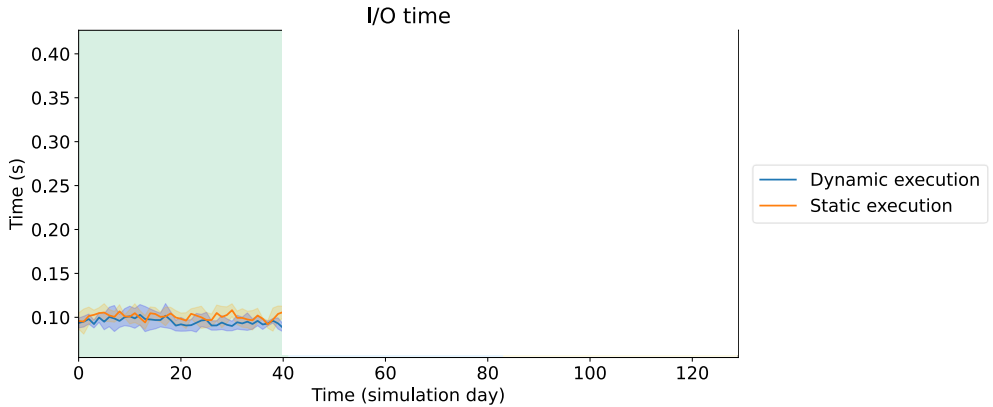
Evaluation

CPU vs I/O time

$$(2,\ 4) \xrightarrow{42} (5,\ 4) \xrightarrow{84} (5,\ 2)$$

	Static execution	Dynamic execution
Execution time	321 s	483 s
Operating time	1.747 s	2.835 s



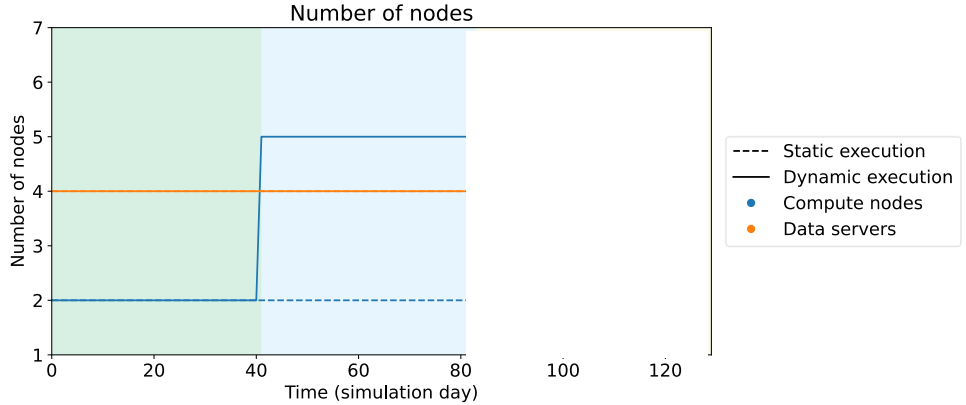


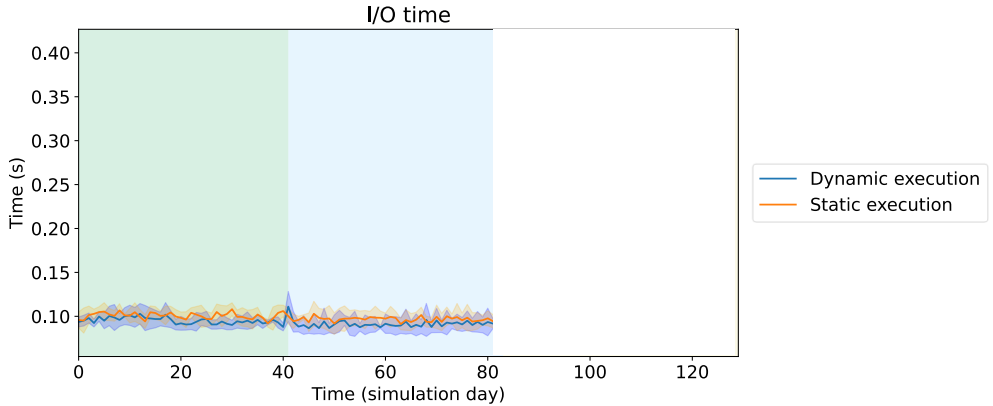
Evaluation

CPU vs I/O time

$$(2,\ 4) \xrightarrow{42} (5,\ 4) \xrightarrow{84} (5,\ 2)$$

	Static execution	Dynamic execution
Execution time	321 s	483 s
Operating time	1.747 s	2.835 s



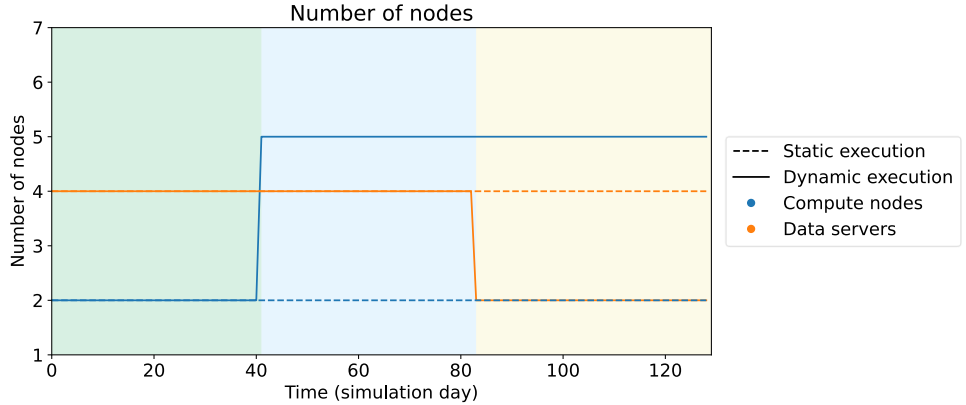


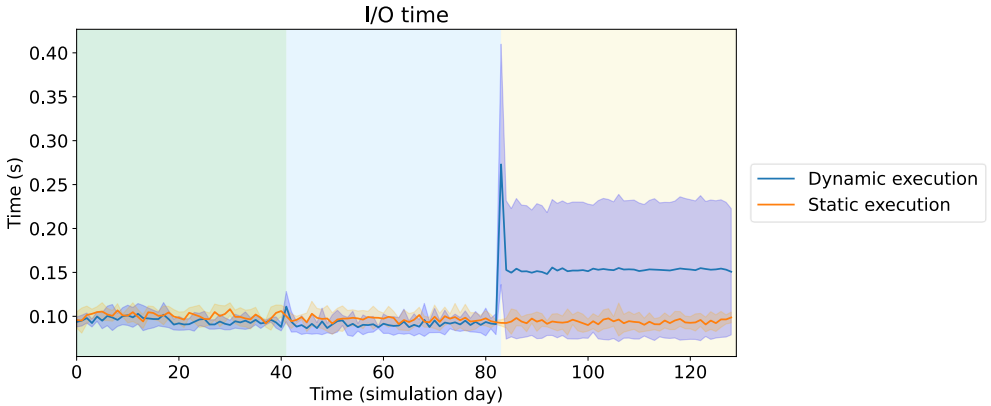
Evaluation

CPU vs I/O time

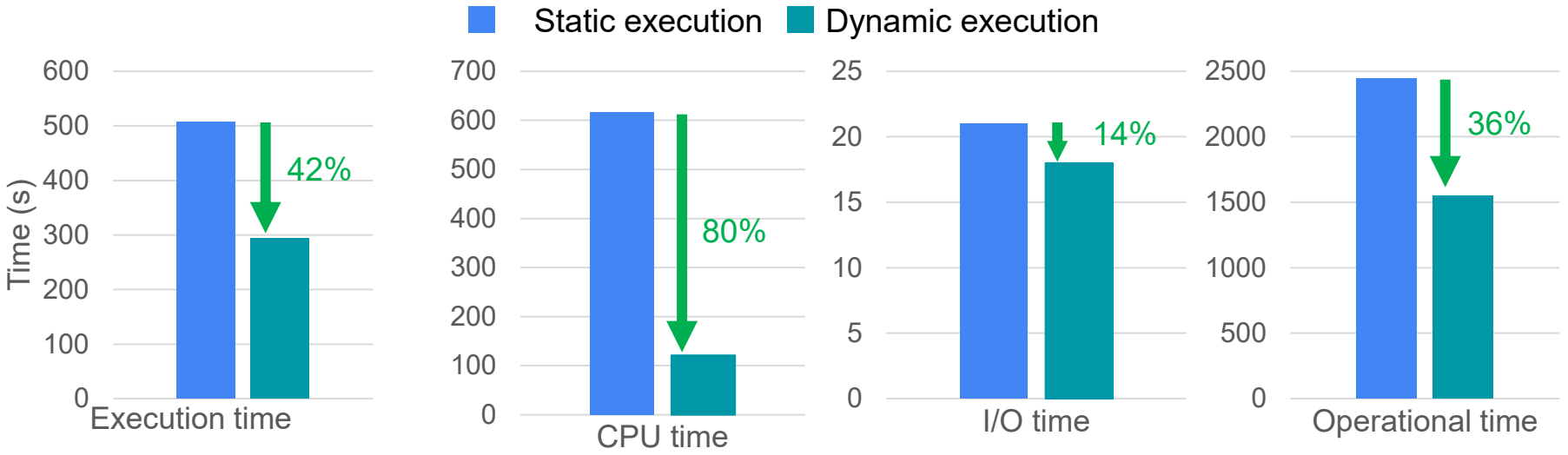
$$(2,\ 4) \xrightarrow{42} (5,\ 4) \xrightarrow{84} (5,\ 2)$$

	Static execution	Dynamic execution
Execution time	321 s	483 s
Operating time	1.747 s	2.835 s





Evaluation


Static execution
Dynamic execution
Time (s)
600
500
400
300
200
100
0
42%
Execution time
700
600
500
400
300
200
100
0
80%
CPU time
25
20
15
10
5
0
14%
I/O time
2500
2000
1500
1000
500
0
36%
Operational time

Conclusion

- Optimizing applications with varying performance is a challenge.
- Dynamic resource management (CPU and I/O) adds additional dimensions to the optimization problem.
- There is significant room for improvement!!
- The approach can be extended to other applications and file systems.