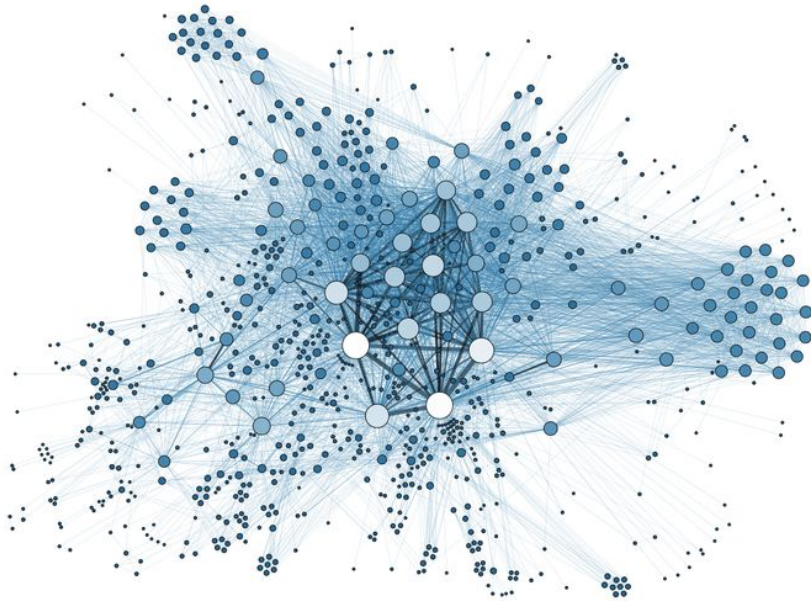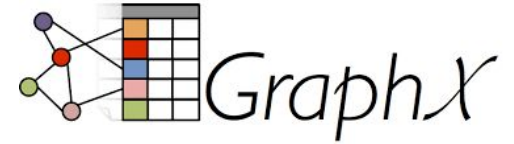# Granula: Toward Fine-grained Performance Analysis of Large-scale Graph Processing Platforms

Wing Lung Ngai, Tim Hegeman, Stijn Heldens,
and Alexandru Iosup

@Large Research
Massivizing Computer Systems

VU UNIVERSITY AMSTERDAM

Large-scale Graph Processing

Towards trillion vertices and edges

Graph processing platforms

Performance Evaluation

Performance evaluation is the process to

- quantify performance,
- explain differences,
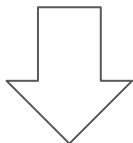- identify overheads
- recommend improvements.

Performance Studies
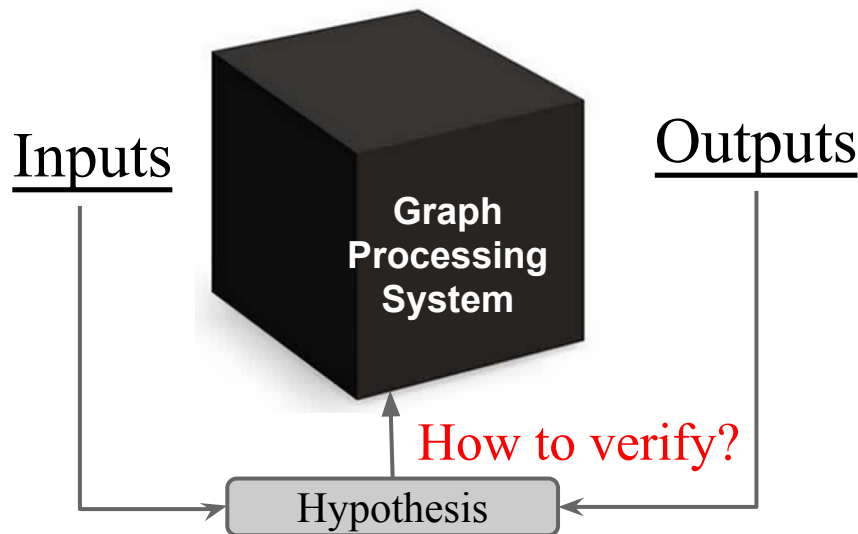

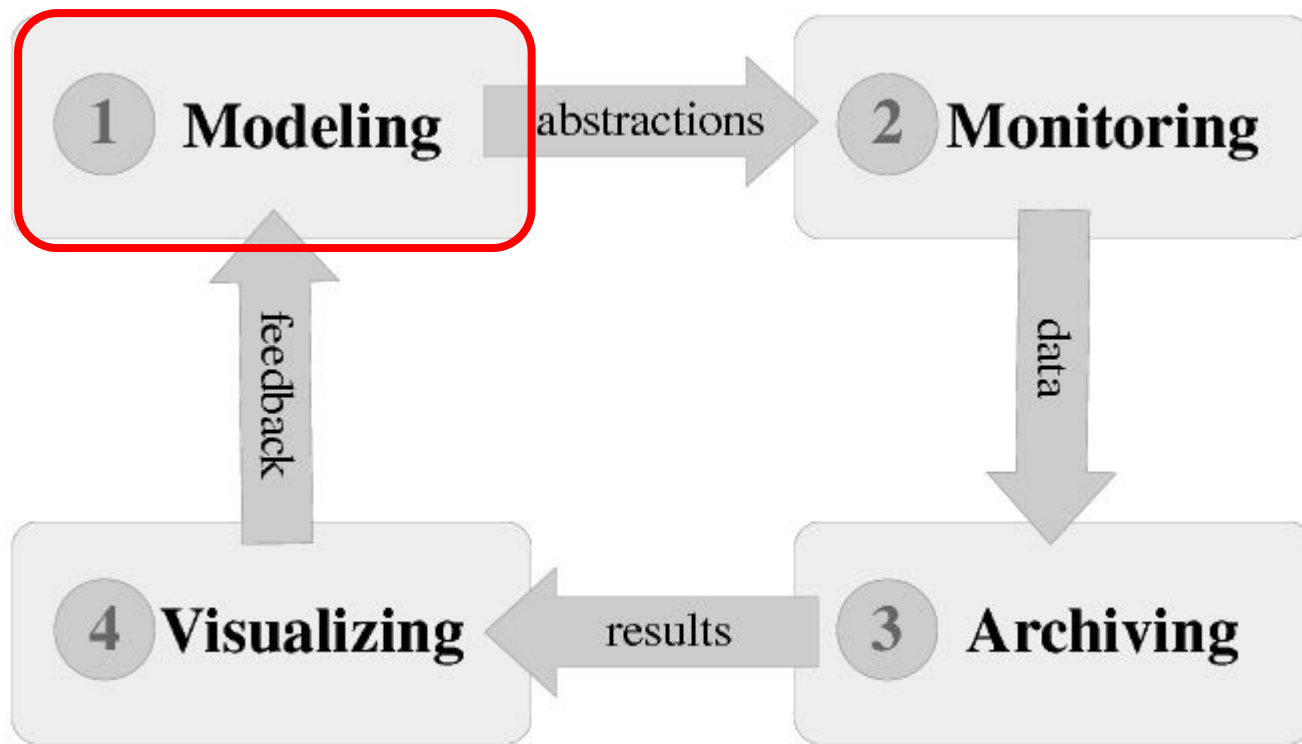User | Developer | Researcher

Identified issues

- coarse-grained vs fine-grained
- limited reusability of studies
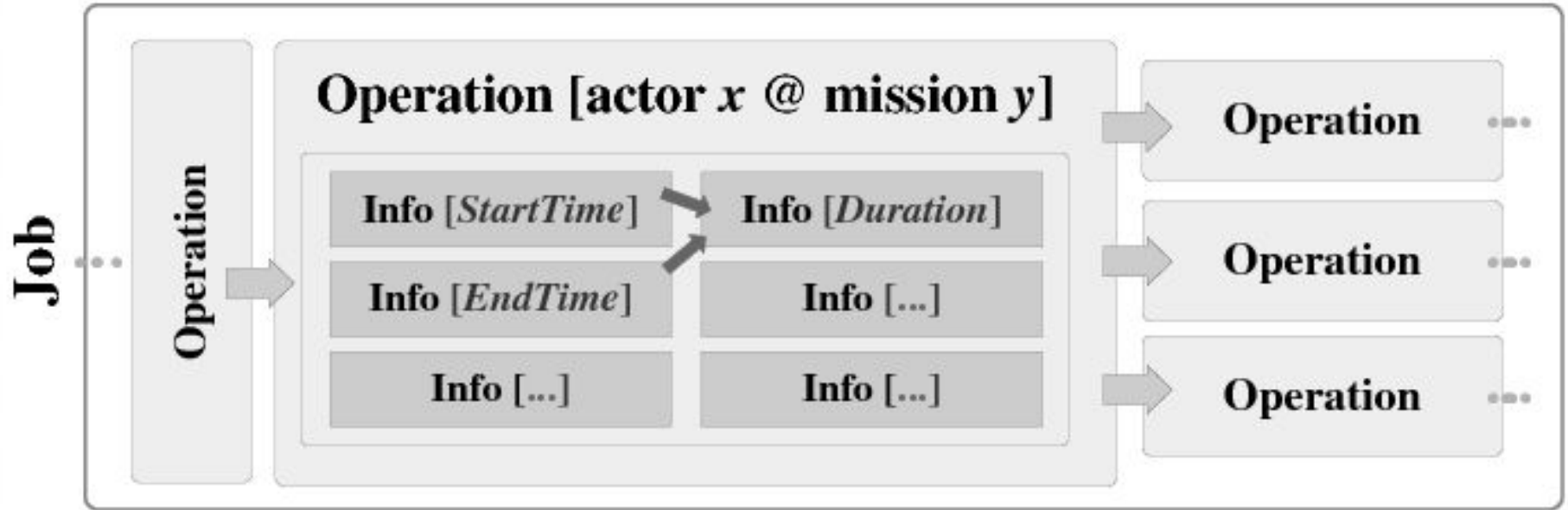- lack of an end-to-end process

Inefficiency of evaluation

Inputs

Outputs

Graph Processing System

How to verify?

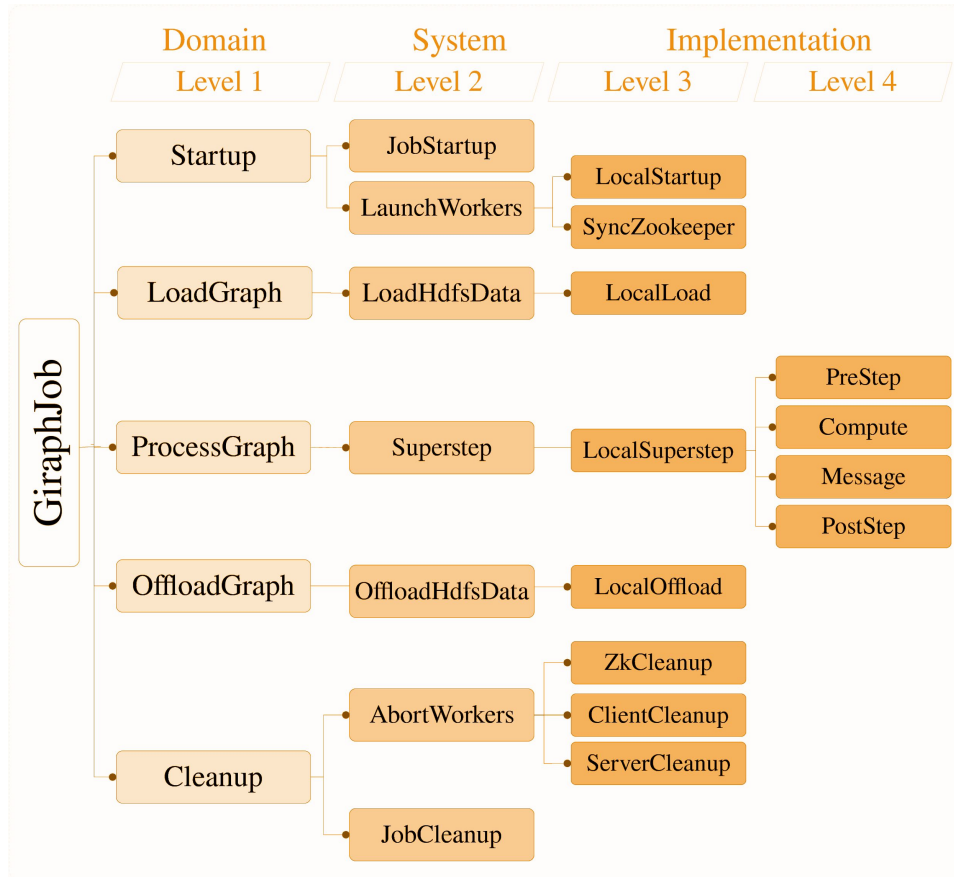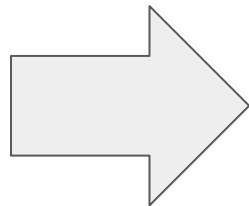Hypothesis

VU UNIVERSITY AMSTERDAM

End-to-end evaluation process

Performance Modeling Language

Building Performance Model

Quantifying System Performance

Running BFS on dg1000 (giraph and powergraph)

Setup	Input/output	Processing

Giraph

0.00%　　20.00%　　40.00%　　60.00%　　80.00%　　100.00%

Startup | LoadGraph | ProcessGraph | Cleanup

0.00s　　16.32s　　32.64s　　48.96s　　65.27s　　81.59s

PowerGraph

0.00%　　20.00%　　40.00%　　60.00%　　80.00%　　100.00%

LoadGraph

0.00s　　80.08s　　160.15s　　240.23s　　320.31s　　400.38s

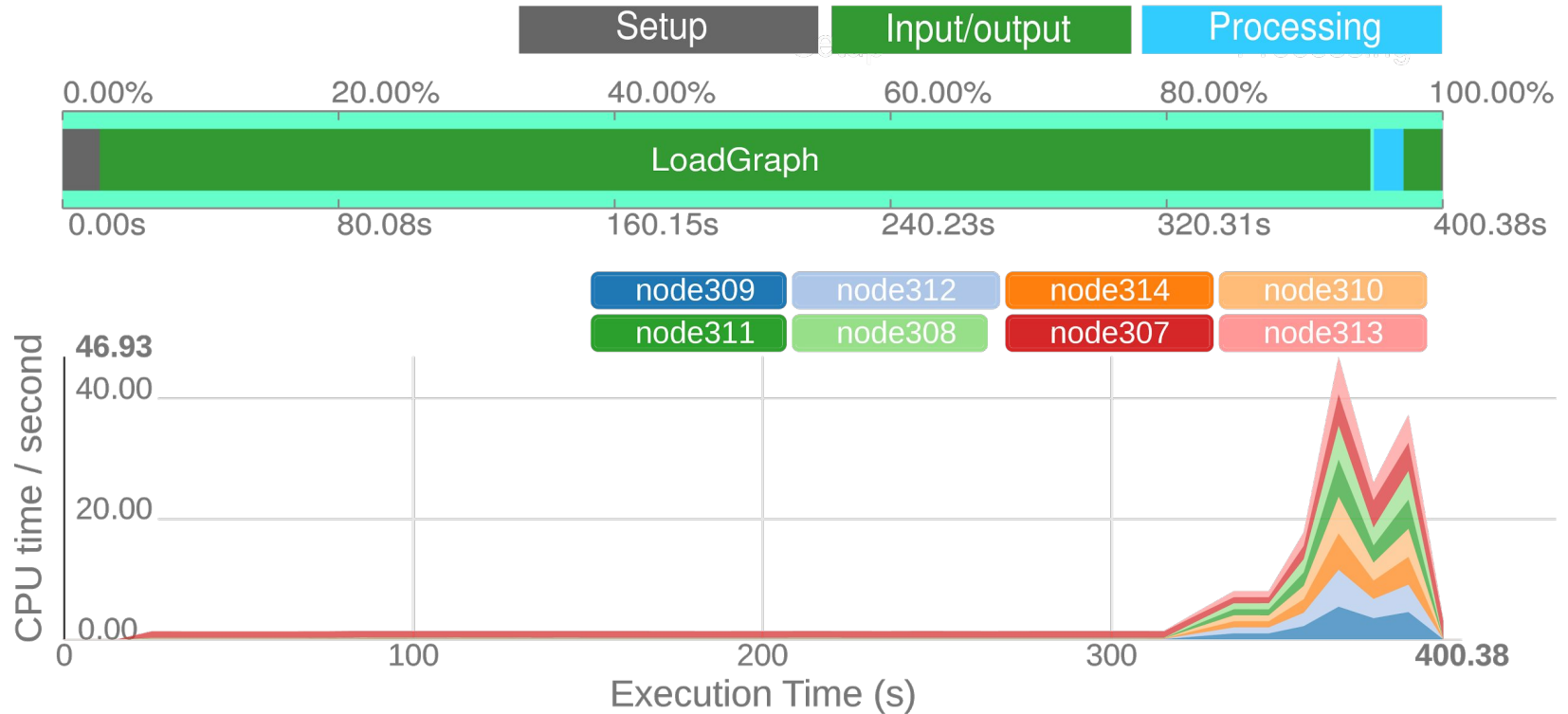VU | UNIVERSITY AMSTERDAM
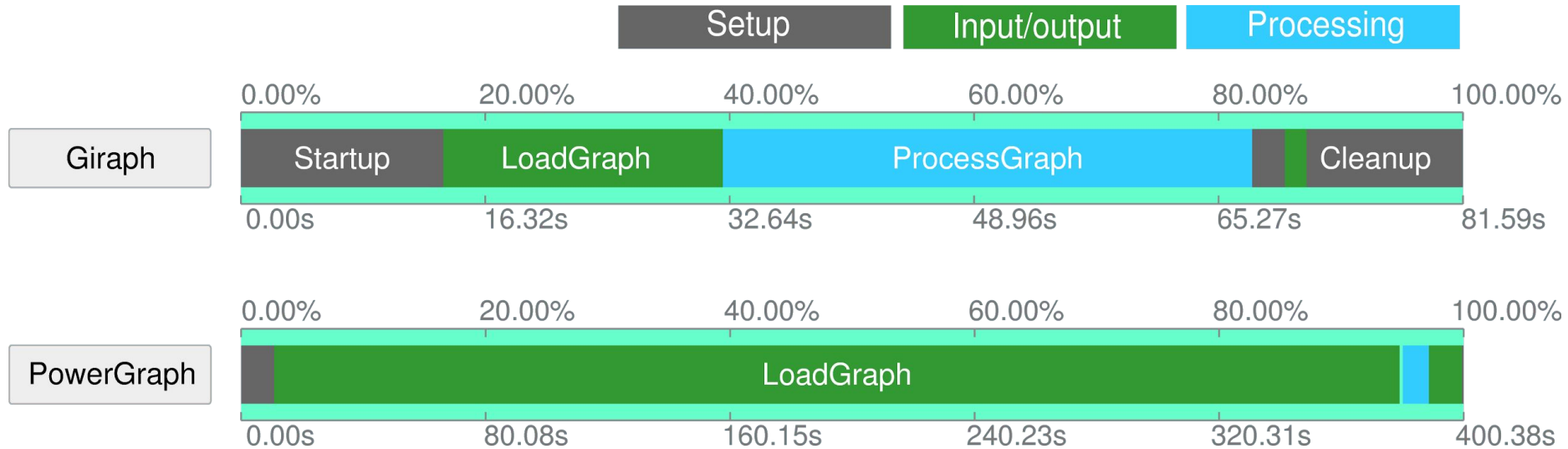
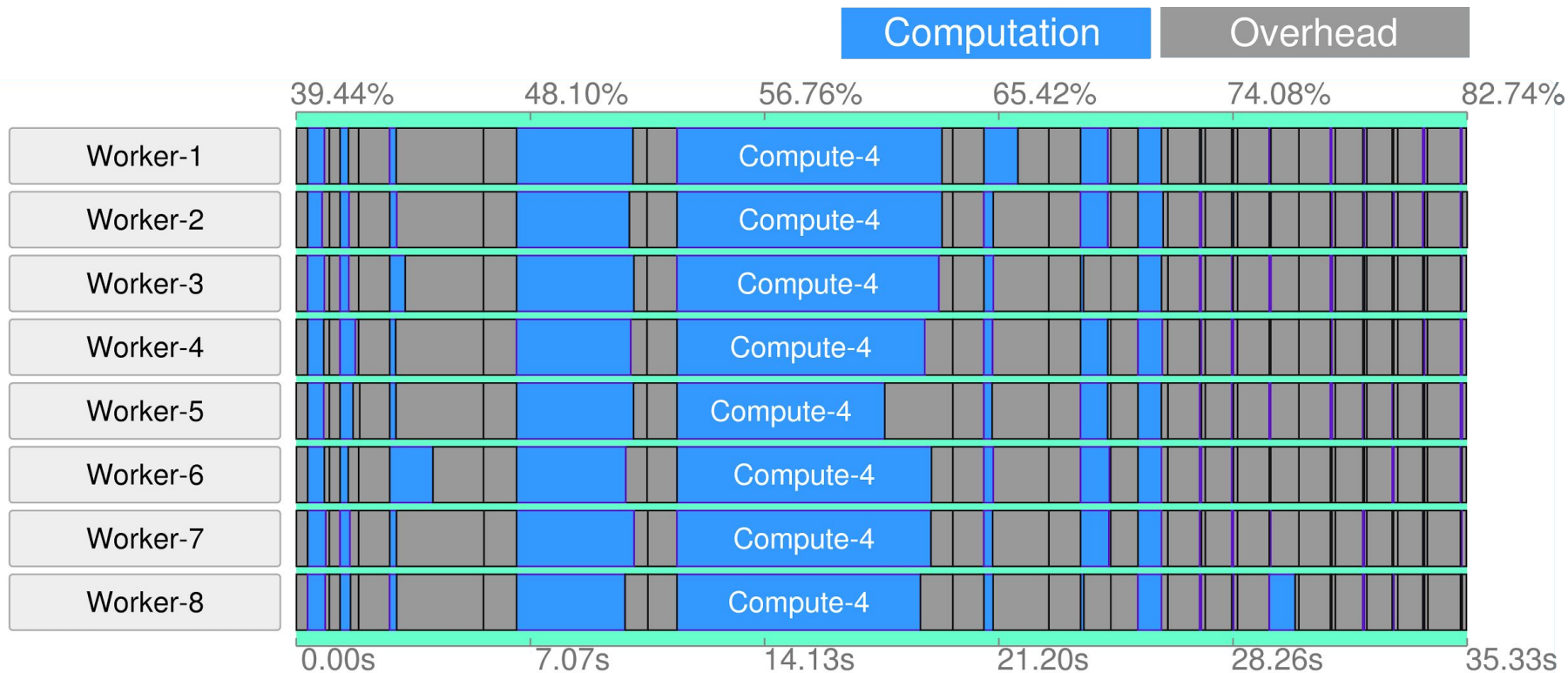# Monitoring Resource Usage

Running BFS on dg1000 (giraph)
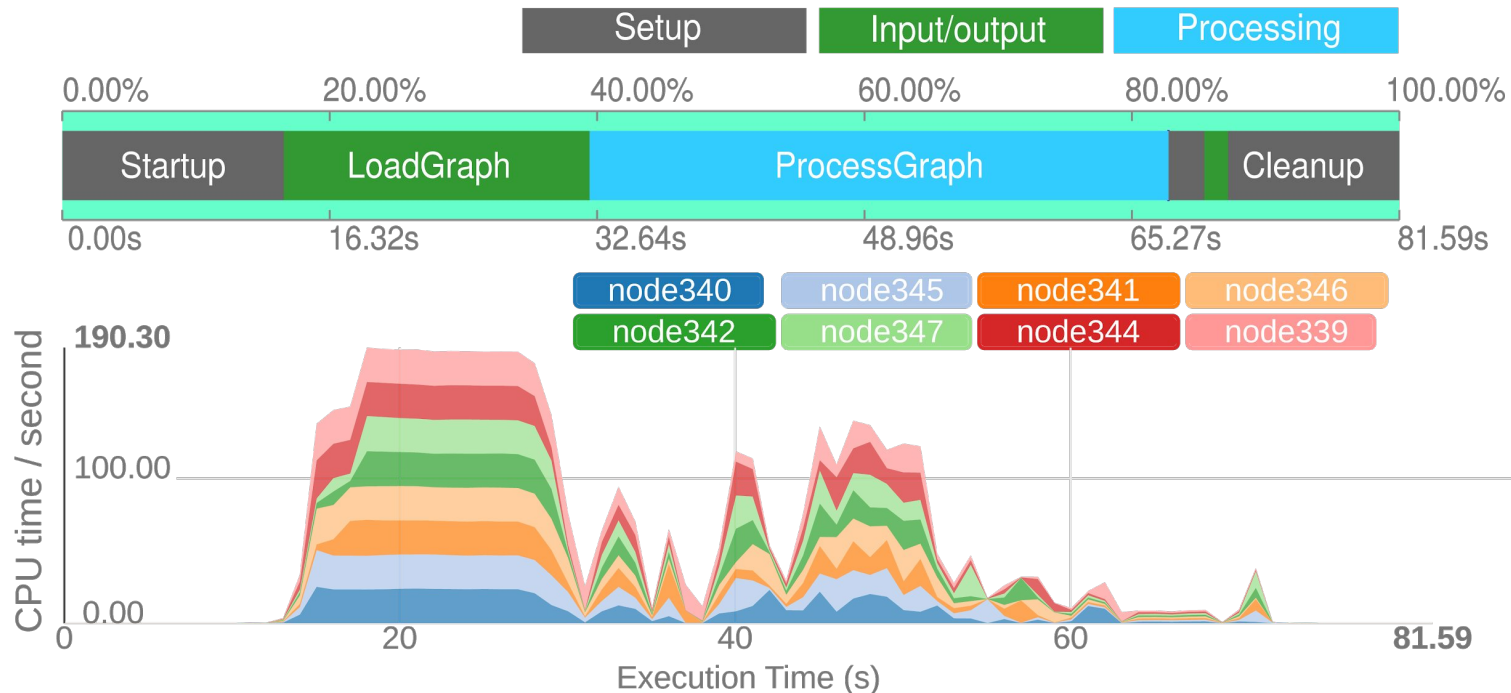
Monitoring Resource Usage

Running BFS on dg1000 (powergraph)

Visualizing System Behavior

Running BFS on dg1000 (giraph)

| Computation | Overhead |

Visualizing System Behavior

⇨ SyncOverhead@[BspMaster @ BspIteration] ⇨ SyncOverhead@[BspWorker-M @ MasterTask]

Target Information

[SyncOverhead] is aggregated from [SyncOverhead]s of all children operation with mission type GlobalSuperstep.

Information [SyncOverhead]   ⌄

Name	Type	Value	Information Owner
SyncOverhead	BasicInfo	1758	Operation [BspWorker-M @ MasterTask]

Source List

Source [SyncOverhead]

Name	Value	Information Owner
SyncOverhead 🔍	132	Operation [GlobalCoordinator @ GlobalSuperstep-0]
SyncOverhead 🔍	45	Operation [GlobalCoordinator @ GlobalSuperstep-1]
SyncOverhead 🔍	25	Operation [GlobalCoordinator @ GlobalSuperstep-2]

Diagnosing Job Failure

Running LCC on dg1000 (powergraph)
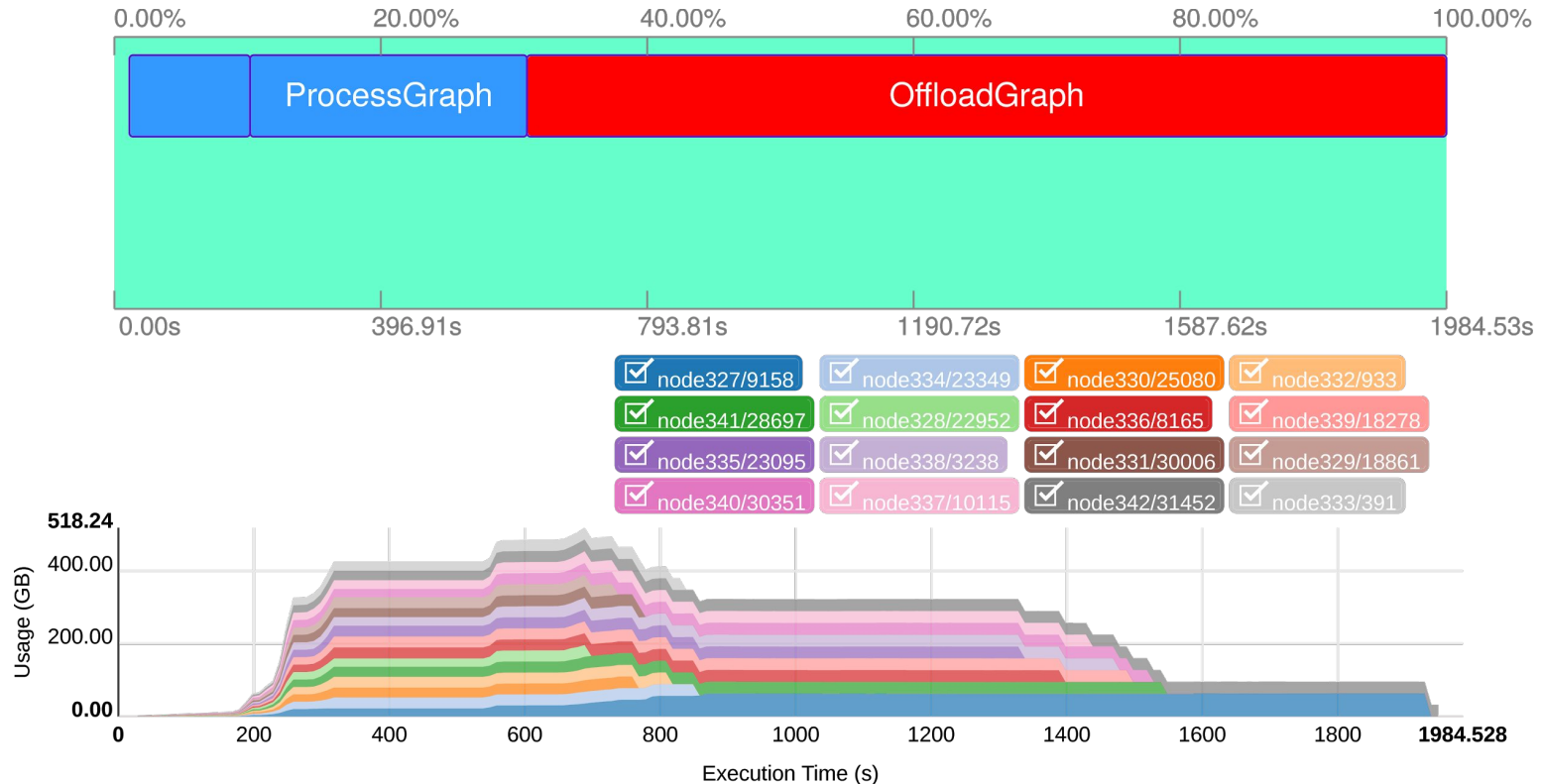
Fine–grained Performance Analysis

- <u>Comprehensive</u>: end-to-end evaluation process.
- <u>Standardized</u>: reusable across platforms.
- <u>Incremental</u>: iterative performance modeling.
- <u>Automatable</u>: automated repetitive procedures.

Granula: a fine-grained performance analysis system for Big Data platforms

VU UNIVERSITY AMSTERDAM

Conclusion

We propose **Granula**:

- a fine-grained performance analysis system for Big Data platforms
- that facilitates modeling, monitoring, archiving, and visualization

Future Work:

- continue the development of our research prototype,
- apply our techniques on other types of Big Data platforms,
- better support for analysts on, e.g, failure diagnosis, regression tests,
- integrate performance analysis into standard software engineering practices

VU UNIVERSITY AMSTERDAM