CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

Towards Measuring and Understanding Performance in Infrastructure- and Function-as-a-Service Clouds

Licentiate Seminar
August 28, 2020, 14:00

Supervisors:
Philipp Leitner
Jan-Philipp Steghöfer

Discussion Leader:
Alessandro Papadopoulos

Joel Scheuner

scheuner@chalmers.se

joe4dev

joelscheuner.com

Supported by WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Methods and insights to guide **performance-optimal cloud service selection**

Evolution of Computing Services

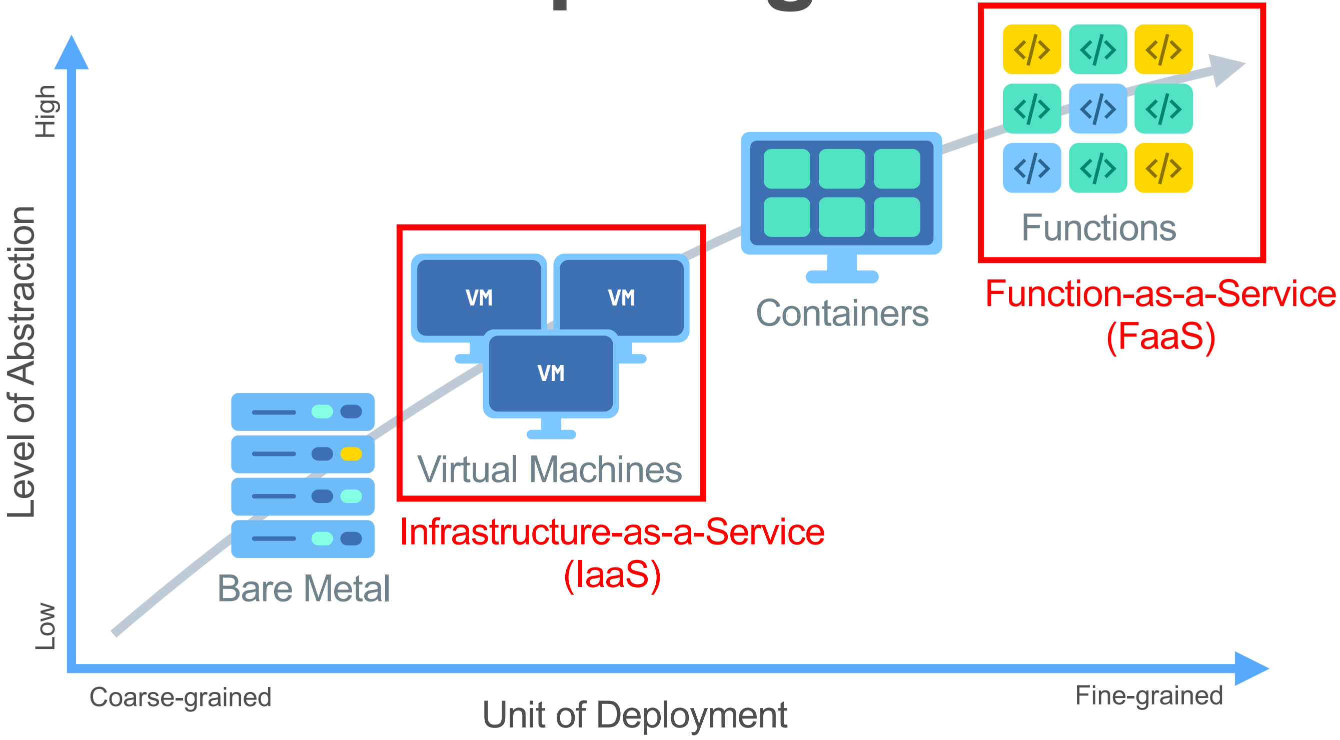
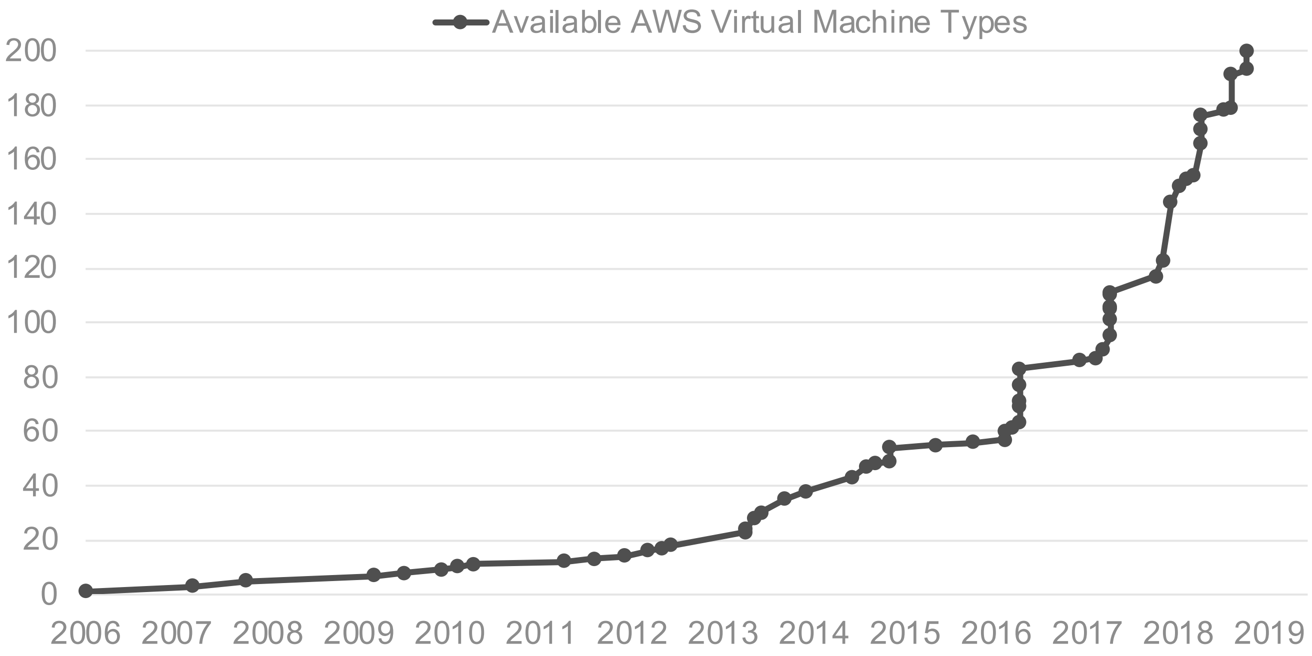
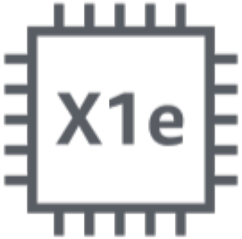


Figure adapted from S. Fink. Serverless – Where Have We Come? Where Are We Going? Keynote at WoSC@CLOUD. 2018.

What cloud service should I choose?



0.05-1 vCPU

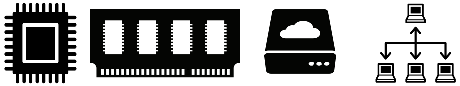


128 vCPUs

Data source: https://aws.amazon.com/blogs/aws/ec2-instance-history/

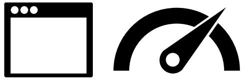
Types of Performance Benchmarks

Micro-Benchmarks



CPU Memory Storage Network

Application-Benchmarks

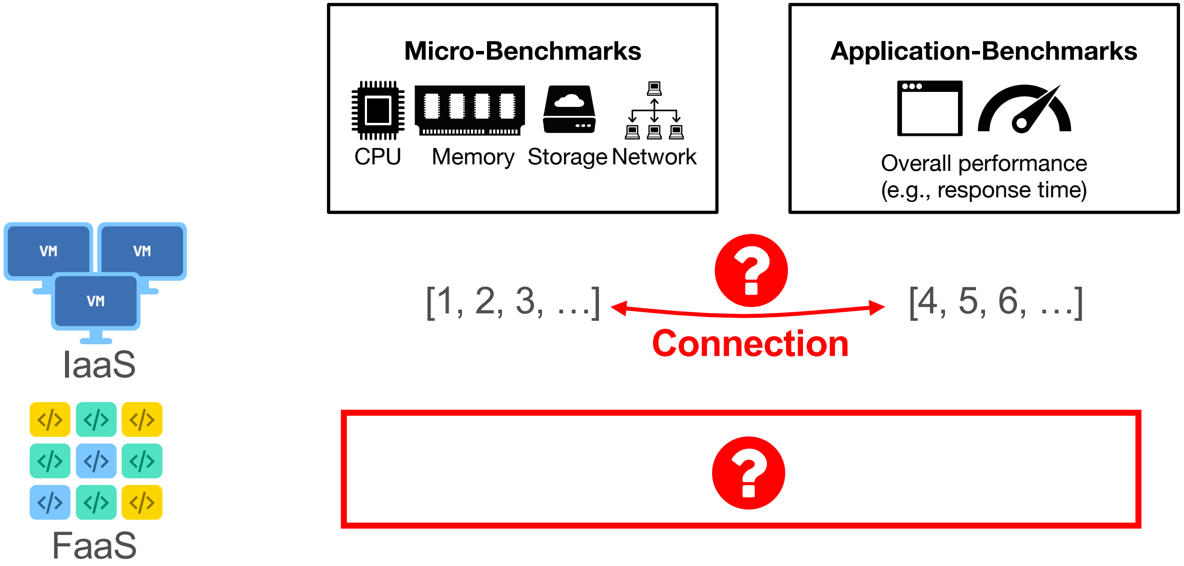


Overall performance
(e.g., response time)

Micro-Benchmarks		Application-Benchmarks
Generic	**Domain**	Specific
Artificial	**Workload**	Real-world
Narrow	**Resource Usage**	Heterogenous

Distinction based on: Z. Li, H. Zhang, L. O'Brien, R. Cai and S. Flint. On Evaluating Commercial Cloud Services: A Systematic Review. Journal of Systems and Software, 2013.

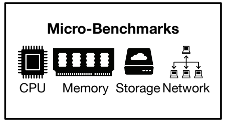
Related Work

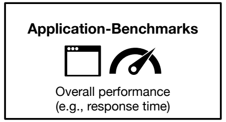


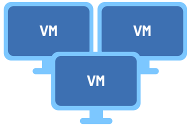
[1] S. Ostermann et al. A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing. Cloud Computing. 2009.
[2] A. Iosup et al. Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing, IEEE Trans. on Parallel and Distributed Systems. 2011.
[3] K. R. Jackson et al. Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud, CloudCom. 2010.
[4] B. F. Cooper et al. Benchmarking Cloud Serving Systems with YCSB, Symposium on Cloud Computing. 2010.
[5] M. Ferdman et al. Clearing the Clouds: A Study of Emerging Scale-out Workloads on Modern Hardware, ASPLOS. 2012.
[6] Y. Gan et al. An Open-Source Benchmark Suite for Microservices and Their Hardware-Software Implications for Cloud & Edge Systems, ASPLOS. 2019.

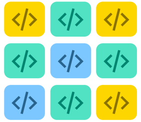
Research Goal

My licentiate thesis aims towards **measuring** and **understanding performance** in **IaaS** and **FaaS** clouds.



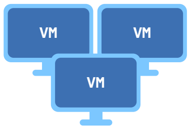






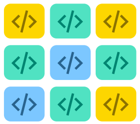
Research Questions

RQ1



How can performance be measured and evaluated in IaaS clouds?

RQ2



What is the current understanding of performance in FaaS clouds?

RQ1: Sub-Questions

RQ1.1: How can multiple performance benchmarks reproducibly evaluate IaaS cloud performance?

RQ1.2: How suitable are micro-benchmarks to estimate application performance in IaaS clouds?

RQ1: Main Findings

RQ1.1: How can multiple performance benchmarks reproducibly evaluate IaaS cloud performance?

Execution methodology combining benchmarks

RQ1.2: How suitable are micro-benchmarks to estimate application performance in IaaS clouds?

Selected micro-benchmarks can be suitable

Benchmarks cannot be used interchangeably

Baseline metrics vCPU and ECU* are insufficient

*provider measure for compute power

RQ1: Research Methodology

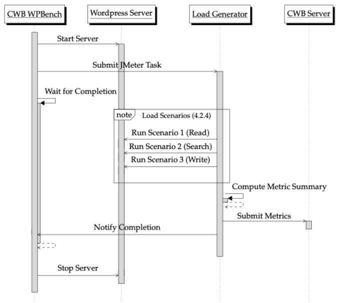
Field Experiment

Benchmark Design → **Benchmark Execution** → **Data Pre-processing** → **Data Analysis**

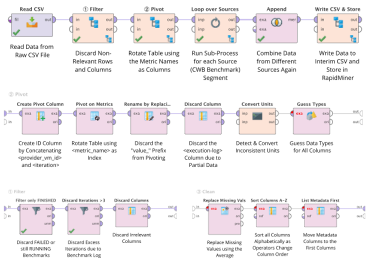
Micro-Benchmarks
- FIO
- StressNg – CPU
- StressNg – Network
- Sysbench – CPU
- Sysbench – File I/O
- Sysbench – Memory
- Sysbench – Threads
- Sysbench – Mutex
- iperf

App-Benchmarks
- MDSim
- WPBench

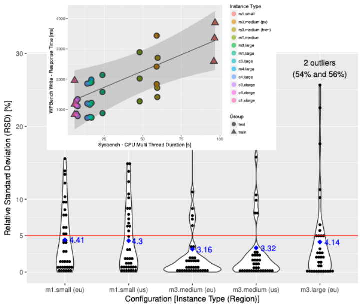
25 Performance Benchmarks



244 Executions



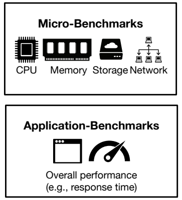
>60 000 Measurements



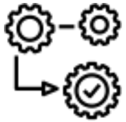
Figures, Tables, and Values

IaaS Benchmark Suite

RQ1.1: How can multiple performance benchmarks reproducibly evaluate IaaS cloud performance?



IaaS benchmark suite
- 23 micro-benchmarks
- 2 application-benchmarks



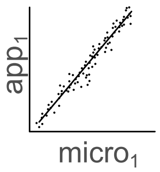
Execution methodology for benchmark suite

Reproducible results* from repeated executions under the same configuration
- 38 benchmark metrics
- 33 executions
- 5 different configurations

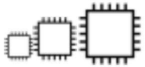
*coefficient of variation <5%

Application Performance Estimation

RQ1.2: How suitable are micro-benchmarks to estimate application performance in IaaS clouds?



Linear regression model



11 Virtual machine types

38 Benchmark metrics

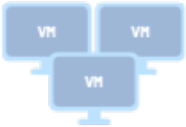
	Web Application (Response Time)	Scientific App. (Duration)
Sysbench CPU Multi-Thread	13%	8%
Sysbench CPU Single-Thread	450%	230%
ECU*	359%	206%

*provider measure for compute power

Relative Error (i.e., MAPE) in percent

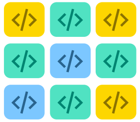
From IaaS to FaaS

RQ1



How can performance be measured and evaluated in IaaS clouds?

RQ2



What is the current understanding of performance in FaaS clouds?

RQ2: Sub-Questions

RQ2.1: What are the characteristics of typical FaaS applications?

RQ2.2: What do existing FaaS performance studies evaluate?

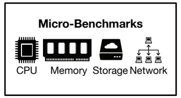
RQ2.3: How reproducible are existing FaaS performance experiments?

RQ2: Main Findings

RQ2.1: What are the characteristics of typical FaaS applications?

FaaS applications typically exhibit workload burstiness

RQ2.2: What do existing FaaS performance studies evaluate?



CPU micro-benchmarks in AWS Lambda are studied most

RQ2.3: How reproducible are existing FaaS performance experiments?

Principles on reproducible cloud experimentation [1] are not followed

Academic studies were not consistently more reproducible

[1] A. V. Papadopoulos, L. Versluis, A. Bauer, N. Herbst, J. von Kistowski, A. Ali-Eldin, C. L. Abad, J. N. Amaral, P. Tuma and A. Iosup. Methodological Principles for Reproducible Performance Evaluation in Cloud Computing. IEEE Transactions on Software Engineering. 2019.

RQ2: Research Methodology

Qualitative Sample Study

89 FaaS applications

24 Characteristics

Documentation and code
→ Primary research

Literature Review

112 FaaS performance studies

51 academic literature

61 grey literature

Studies and their design
→ Secondary research

FaaS Applications

RQ2.1: What are the characteristics of typical FaaS applications?

100% 89 FaaS applications

Workload Burstiness

- 84% Yes
- 26% No

FaaS Motivators*

- 47% Save costs
- 34% Built-in scalability
- 34% No operations

Latency Requirements

- 38% None
- 32% Complete application
- 28% Parts of the application

* Unknown for 30% of applications. Detailed results in accompanying technical report S. Eismann, J. Scheuner, E. van Eyk, M. Schwinger, J. Grohmann, N. Herbst, C. L. Abad, and A. Iosup A Review of Serverless Use Cases and their Characteristics, SPEC RG Cloud Working Group. 2020.

Existing FaaS Performance Studies

RQ2.2: What do existing FaaS performance studies evaluate?

Literature Type

100% 51 academic literature studies
100% 61 grey literature studies

Programming Languages

42% JavaScript
42% Python
12% Java

Benchmark Type

67% 82%	Micro-benchmarks
57% 31%	Application-benchmarks
24% 13%	Both

Deployment Platform

88% 89%	AWS Lambda
27% 25%	Azure Functions
27% 20%	Google Cloud Functions

Language Runtimes

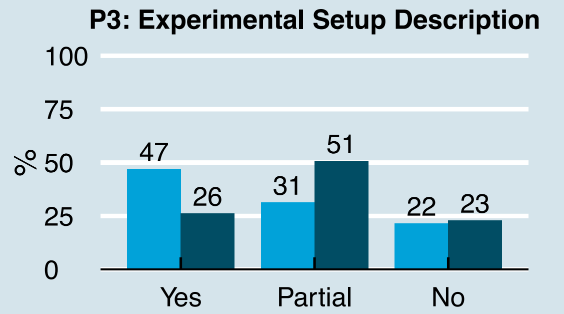
49% 43%	Python
29% 56%	Node.js
14% 34%	Java

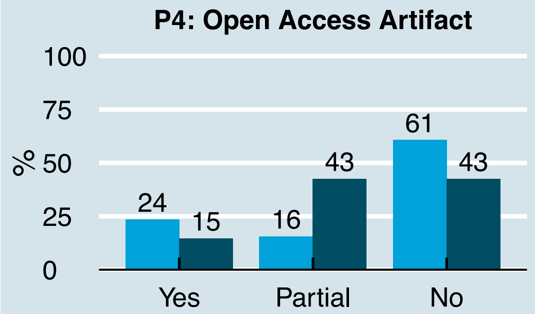
Reproducibility of FaaS Experiments

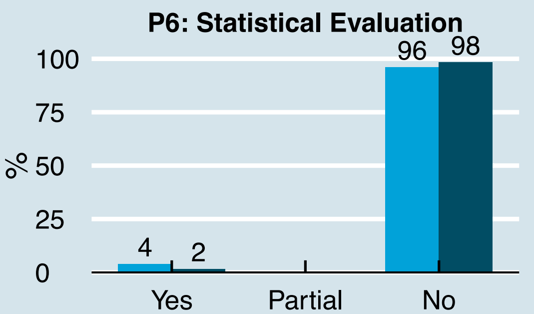
RQ2.3: How reproducible are existing FaaS performance experiments?

Following existing principles on reproducible cloud experimentation [1]

Literature Type: Academic, Grey







[1] A. V. Papadopoulos, L. Versluis, A. Bauer, N. Herbst, J. von Kistowski, A. Ali-Eldin, C. L. Abad, J. N. Amaral, P. Tuma and A. Iosup. Methodological Principles for Reproducible Performance Evaluation in Cloud Computing. IEEE Transactions on Software Engineering. 2019.

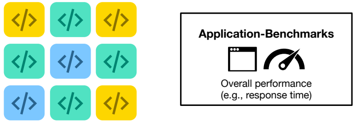
Conclusion

Improve future cloud performance evaluation studies

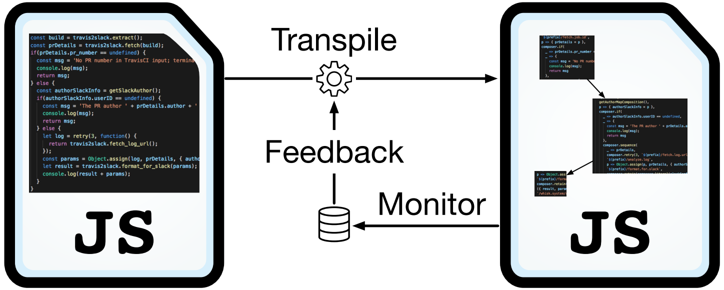
Guide performance-optimal cloud service selection

Ongoing Work

1) FaaS application performance benchmark



2) Performance-optimized FaaS applications



Application Code

FaaS Orchestration

Joel Scheuner

scheuner@chalmers.se

joe4dev

joelscheuner.com

Credits

- Icons made by Freepik, monkik, geotatah, mynamepong, and smashicons from www.flaticon.com