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
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?

Available AWS Virtual Machine Types

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)

**Domain**

- Generic
- Artificial
- Narrow

**Workload**

- Specific
- Real-world

**Resource Usage**

- Heterogenous

Related Work

Micro-Benchmarks
- CPU
- Memory
- Storage
- Network

Application-Benchmarks
- Overall performance (e.g., response time)

[1, 2, 3, …] Connection [4, 5, 6, …]

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

- Benchmark Design
  - Micro-Benchmarks
    - FIO
    - StressNg – CPU
    - StressNg – Network
    - Sysbench – CPU
    - Sysbench – File I/O
    - Sysbench – Memory
    - Sysbench – Threads
    - Sysbench – Mutex
    - iperf
  - App-Benchmarks
    - MDSim
    - WPBench

- Benchmark Execution
  - 25 Performance Benchmarks
  - 244 Executions

- Data Pre-processing
  - >60 000 Measurements

- Data Analysis
  - 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

*coefficient of variation <5%

- 38 benchmark metrics
- 33 executions
- 5 different configurations
**Application Performance Estimation**

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

<table>
<thead>
<tr>
<th></th>
<th>Web Application (Response Time)</th>
<th>Scientific App. (Duration)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sysbench CPU Multi-Thread</td>
<td>13%</td>
<td>8%</td>
</tr>
<tr>
<td>Sysbench CPU Single-Thread</td>
<td>450%</td>
<td>230%</td>
</tr>
<tr>
<td>ECU*</td>
<td>359%</td>
<td>206%</td>
</tr>
</tbody>
</table>

*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

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

<table>
<thead>
<tr>
<th>Workload Burstiness</th>
<th>FaaS Motivators*</th>
<th>Latency Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>84% Yes</td>
<td>47% Save costs</td>
<td>38% None</td>
</tr>
<tr>
<td>26% No</td>
<td>34% Built-in scalability</td>
<td>32% Complete application</td>
</tr>
<tr>
<td></td>
<td>34% No operations</td>
<td>28% Parts of the application</td>
</tr>
</tbody>
</table>

Existing FaaS Performance Studies

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

<table>
<thead>
<tr>
<th>Literature Type</th>
<th>100% 51 academic literature studies</th>
<th>100% 61 grey literature studies</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Benchmark Type</th>
<th>67% Micro-benchmarks</th>
<th>82% Application-benchmarks</th>
<th>57% Both</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Deployment Platform</th>
<th>88% AWS Lambda</th>
<th>89% Azure Functions</th>
<th>27% Google Cloud Functions</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Programming Languages</th>
<th>42% JavaScript</th>
<th>42% Python</th>
<th>12% Java</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Language Runtimes</th>
<th>49% Python</th>
<th>43% Node.js</th>
<th>29% Java</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Workload Burstiness</th>
<th>80% AWS Lambda</th>
<th>10% Azure Functions</th>
<th>34% Google Cloud Functions</th>
</tr>
</thead>
</table>
Reproducibility of FaaS Experiments

RQ2.3: How reproducible are existing FaaS performance experiments? Following existing principles on reproducible cloud experimentation [1]

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

Joel Scheuner
✉ scheuner@chalmers.se
👋 joe4dev
🔗 joelscheuner.com
Credits

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