Towards a Resource Elasticity Benchmark for Cloud Environments

Andreas Weber, Nikolas Herbst, Henning Groenda, Samuel Kounev

HotTopics’14, Dublin, Ireland - March 22nd, 2014
How would you define elasticity?

**OCDA** [OCDA12]  
up & down scaling  
subsriber workload

**IBM, Schouten**  
[Schouten12]  
scalability  
increase & reduce  
no manual labor

**Cohen** [Cohen09]  
quantifyable  
real-time demands  
local & remote

**NIST** [Mell11]  
rrapid elasticity unlimited  
provission & release  
sometimes automated  
with demand

**Eukalyptus, Wolski**  
[Wolski11]  
measurable  
mapping of  
requests to resources

<table>
<thead>
<tr>
<th>What is Elasticity?</th>
<th>Motivation</th>
<th>Related Work</th>
<th>Approach</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weber, Herbst, Groenda, Kounnev</td>
<td>Towards a Resource Elasticity Benchmark in Cloud Environments</td>
<td></td>
<td>HotTopics’14, Dublin, Ireland - March 22nd</td>
<td></td>
</tr>
</tbody>
</table>
### What is Elasticity and how can we measure it?

<table>
<thead>
<tr>
<th>What is Elasticity?</th>
<th>Motivation</th>
<th>Related Work</th>
<th>Approach</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weber, Herbst, Groenda, Kounev</td>
<td>Towards a Resource Elasticity Benchmark in Cloud Environments</td>
<td>HotTopics’14, Dublin, Ireland - March 22nd</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is Elasticity and how can we measure it?

- **What is Elasticity?**
- **Motivation**
- **Related Work**
- **Approach**
- **Conclusion**

<table>
<thead>
<tr>
<th>What is Elasticity?</th>
<th>Motivation</th>
<th>Related Work</th>
<th>Approach</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weber, Herbst, Groenda, Kounev</td>
<td>Towards a Resource Elasticity Benchmark in Cloud Environments</td>
<td></td>
<td></td>
<td>HotTopics’14, Dublin, Ireland - March 22nd</td>
</tr>
</tbody>
</table>
What is Elasticity and how can we measure it?

Motivation

Related Work

Approach

Conclusion

Weber, Herbst, Groenda, Kounov

Towards a Resource Elasticity Benchmark in Cloud Environments
What is Elasticity and how can we measure it?

- What is Elasticity?

- Related Work
  - Weber, Herbst, Groenda, Kounev

- Approach

- Conclusion
Motivation

Elasticity:
- Mayor quality attribute of clouds [Gartner09]
- Many strategies exist [Galante12, Jennings14]
- Industry & Academia

→ Benchmark for comparability!

What is Elasticity? | Motivation | Related Work | Approach | Conclusion
---|---|---|---|---
Related Work: Elasticity Benchmarking Approaches

- **Specialized approaches**
  - Measure technical provisioning time
  - Measure SLA compliance
  - Focus on scale up/out

- **Business perspective**
  - What is the financial impact?
  - Disadvantage: Mix-up of elasticity technique and business model

[Binning09, Li10, Dory11, Almeida13, Islam12, Folkerts12]
Cloud System Under Test

System Under Test (SUT)

Cloud Management

Elasticity Mechanism

Monitoring System

Reconfiguration Management

Load Balancer

Elastic Infrastructure

Active VMs

Hypervisor

Host

1
2
3
4
5
6
...
...
...

What is Elasticity?

Motivation

Related Work

Approach

Conclusion

Weber, Herbst, Groenda, Kounev

Towards a Resource Elasticity Benchmark in Cloud Environments

HotTopics’14, Dublin, Ireland - March 22nd
Approach - Workflow

**Calibration**
- Analyze Scalability
- Adjust Load Profile

**Measurement**
- Penetrate SUT
- Monitor Resource Supply
- Extract Demand & Supply
- Apply Metrics

The process flow starts with **Calibration**, followed by **Measurement**.

**What is Elasticity?**

**Motivation**

**Related Work**

**Approach**

**Conclusion**

Weber, Herbst, Groenda, Kounev

Towards a Resource Elasticity Benchmark in Cloud Environments

HotTopics’14, Dublin, Ireland - March 22nd
Prerequisite: Service Level Objective (SLO)

Goal:
Find a mapping between
- load intensity
- resource demand [\# resources]

Service Level Objective (SLO):
E.g.: resp. time $\leq$ 1 sec, 95%
Analyze Scalability

**Approach:**
- Disable autoscaling
- Evaluate system separately at each scale
- Find maximal intensity that the system can withstand without violating SLO (binary search)
- Derive demand step function: \( \text{resourceDemand} = f(\text{intensity}) \)

**Benefit:**
- Derive resource demand for arbitrary load intensity variations

What is Elasticity? | Motivation | Related Work | Approach | Conclusion
---|---|---|---|---
Weber, Herbst, Groenda, Kounev | Towards a Resource Elasticity Benchmark in Cloud Environments | HotTopics’14, Dublin, Ireland - March 22nd
Towards a Resource Elasticity Benchmark in Cloud Environments

**Approach - Workflow**

**Calibration**
- Analyze Scalability
- Adjust Load Profile

**Measurement**
- Penetrate SUT
- Monitor Resource Supply
- Extract Demand & Supply
- Apply Metrics

---

What is Elasticity?  |  Motivation  |  Related Work  |  Approach  |  Conclusion
---|---|---|---|---
Weber, Herbst, Groenda, Kounev  |  Towards a Resource Elasticity Benchmark in Cloud Environments  |  |  | HotTopics’14, Dublin, Ireland - March 22nd
Benchmark Calibration

- Same load intensity variation, different efficiency

![Graph showing demand, supply, and time with different profiles for comparison difficulty.]

- Different demand → Comparison difficult

- Induce same resource demand on all systems
  - Analyze scaling behavior
  - Adjust load profile

What is Elasticity? | Motivation | Related Work | Approach | Conclusion
---|---|---|---|---
Weber, Herbst, Groenda, Kounev | Towards a Resource Elasticity Benchmark in Cloud Environments | | HotTopics’14, Dublin, Ireland - March 22nd


Towards a Resource Elasticity Benchmark in Cloud Environments

**Approach - Workflow**

**Calibration**
- Analyze Scalability
- Adjust Load Profile

**Measurement**
- Penetrate SUT
- Monitor Resource Supply
- Extract Demand & Supply
- Apply Metrics

---

**What is Elasticity?**

**Motivation**

**Related Work**

**Approach**
- Weber, Herbst, Groenda, Kounev

**Conclusion**

HotTopics’14, Dublin, Ireland - March 22"
Penetrate SUT

- Challenge: Stress SUT in a representative manner
  - Realistic variability of load intensity
  - Adaptability of load profiles to suit different domains

- Approach:
  - Open workload model [Schroeder06]
  - Model Load Variations with the LIMBO toolkit [Kistowski14]
    - Facilitates creation of new load profiles
      - Derived from existing traces
      - With desired properties (e.g. seasonal pattern, bursts)
  - Execute load profile using JMeter
    - Customized Timer delays requests according to timestamp file created by LIMBO
**Approach - Workflow**

**Calibration**
- Analyze Scalability
- Adjust Load Profile

**Measurement**
- Penetrate SUT
- Monitor Resource Supply
- Extract Demand & Supply
- Apply Metrics

---

**What is Elasticity?**

**Motivation**

**Related Work**

**Approach**

**Conclusion**

17 Weber, Herbst, Groenda, Kounev

Towards a Resource Elasticity Benchmark in Cloud Environments

HotTopics’14, Dublin, Ireland - March 22nd
Approach - Workflow

**Calibration**
- Analyze Scalability
- Adjust Load Profile

**Measurement**
- Penetrate SUT
- Monitor Resource Supply
- Extract Demand & Supply
- Apply Metrics

---

**What is Elasticity?**

**Motivation**

**Related Work**

**Approach**

**Conclusion**
Approach - Workflow

 Calibration

 Analyze Scalability  Adjust Load Profile

 Measurement

 Penetrate SUT  Extract Demand & Supply  Apply Metrics

 Monitor Resource Supply

 What is Elasticity?  Motivation  Related Work  Approach  Conclusion

 Weber, Herbst, Groenda, Kounev

 Towards a Resource Elasticity Benchmark in Cloud Environments

 HotTopics’14, Dublin, Ireland - March 22nd
Measurement – Metric Ideas

- **Accuracy**: $\frac{\sum U}{T}, \frac{\sum O}{T}$ [Herbst13]
- **Timing**: $\frac{|D_u - A_u|}{D_u}$
- **DTW distance** [Kuperberg12]

What is Elasticity?  
Motivation  
Related Work  
Approach  
Conclusion

- $D_u$: #scale up events - demand  
- $A_u$: #scale up events - supply

- Weber, Herbst, Groenda, Kounev  
Towards a Resource Elasticity Benchmark in Cloud Environments
Approach – Workflow

Calibration
- Analyze Scalability
- Adjust Load Profile

Measurement
- Penetrate SUT
- Monitor Resource Supply
- Extract Demand & Supply
- Apply Metrics
## Conclusion

**Motivation**
- Current elasticity rating approaches
  - specialized
  - mix business and technical aspects

**Approach**
- Variable load profiles
- Calibration for equal demand curves
- Comparison of resource demand and allocation with metrics

**Contribution**
- Prototype for a benchmarking framework
  - Rating of elasticity core aspects
  - Independent of cost model and hardware efficiency

**Next Steps**
- Refinement of benchmarking framework
- Evaluation and refinement of metrics
Questions?

- Other important elasticity aspects?
- Global metric or metrics targeted at aspects?

weber | groenda @ fzi.de
herbst | kounev @ kit.edu

www.descartes-research.net
Towards a Resource Elasticity Benchmark in Cloud Environments

**Literature (1/2)**

- **OCDA12:**

- **Mell11:**

- **Shouten12:**

- **Wolski11:**

- **Cohen09:**

- **Gartner09:**

- **Galante12:**

- **Jennings14:**
Binning09:

Li10:

Dory11:

Almeida13:

Islam12:

Folkerts12:

Schroeder06:

Kistowski14:

Herbst13:

Kuperberg12: