IN4392 Cloud Computing
Introduction to Cloud Computing

Alexandru Iosup
Parallel and Distributed Systems Group
Delft University of Technology
The Netherlands

Our team: **Undergrad** Thomas de Ruiter, Anand Sawant, Ruben Verboon, ...
**Grad** Siqi Shen, Guo Yong, Nezih Yigitbasi **Staff** Henk Sips, Dick Epema, Alexandru Iosup, Otto Visser **Collaborators** Ion Stoica and the Mesos team (UC Berkeley), Thomas Fahringer, Radu Prodan, Vlad Nae (U. Innsbruck), Nicolae Tapus, Mihaela Balint, Vlad Posea (UPB), Derrick Kondo, Emmanuel Jeannot (INRIA), Assaf Schuster, Mark Silberstein, Orna Ben-Yehuda (Technion), ...

September 12, 2012
IN4392 Cloud Computing
What is Cloud Computing?
1. A Cloudy Buzzword

- 18 definitions in computer science (ECIS’10). NIST has one. Cal has one. We have one.
- “We have redefined cloud computing to include everything that we already do.” Larry Ellison, Oracle, 2009

What is Cloud Computing?
2. A Descendant* of the Grid Idea

“A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities [+ for] nontrivial QoS.” I. Foster, 1998 + 1999

* Subset.


Grid Stack

Cloud
Grid Applications

Cloud
Grid Very High Level MW

Cloud
Grid High Level MW

Cloud
Grid Low Level MW

Virtualized HW + OS

MW = Middleware
What is Cloud Computing?

3. A Useful IT Service

“Use only when you want! Pay only for what you use!”

- **Software as a Service (SaaS)**
  - Q: What do you use?
  - Q: Why not this level?

- **Platform as a Service (PaaS)**
  - Q: Why not this level?

- **Infrastructure as a Service (IaaS)**
  - Processing Resources
  - Storage Resources
  - Network Resources

September 12, 2012
Agenda

1. What is Cloud Computing?
2. IaaS Clouds, the Core Idea
3. The IaaS Owner Perspective
4. The IaaS User Perspective
5. Reality Check
6. Conclusion
IaaS Cloud Computing
Joe Has an Idea ($$$)

MusicWave

Solution #1
Buy or Rent

- Big up-front commitment

- Load variability: NOT supported

Solution #2
Deploy on IaaS Cloud

- NO big up-front commitment
- Load variability: supported

Q: So are we just shifting the problem to somebody else, that is, the IaaS cloud owner?

Inside an IaaS Cloud Data Center

Time and Cost Sharing Among Users

(Source: A. Antoniou, MSc Defense, TU Delft, 2012.)
Main Characteristics of IaaS Clouds

1. On-Demand Pay-per-Use
2. Elasticity (cloud concept of Scalability)
3. Resource Pooling
4. Fully automated IT services
5. Quality of Service
Agenda

1. What is Cloud Computing?
2. IaaS Clouds, the Core Idea
3. **The IaaS Owner Perspective: How to Deploy a Cloud?**
4. The IaaS User Perspective
5. Reality Check
6. Conclusion
IaaS Cloud Deployment Models

Private
On-premises

Public
Off-premises

Hybrid

Resource Sharing Models

Grids
Space-Sharing

MusicWave
OtherApp
Host OS

Q: Which one is better?

IaaS Clouds
Time-Sharing

MusicWave
OtherApp
Host OS

OtherApp

Virtualization

**Q: What is the problem?**

**Q: What to do now?**
Virtualization and The Full IaaS Stack

- Applications
- Guest OS
- Virtual Resources
- VM Instance
- Virtual Machine Manager
- Virtual Infrastructure Manager
- Physical Infrastructure

- Storage
- Network
- Servers
- Storage
The Virtual Machine Lifecycle

1. Requested
2. Pending
3. Booting
4. Running
5. Shutting-down
6. Terminated

Q: Is this fair?

(Source: A. Antoniou, MSc Defense, TU Delft, 2012.)
Use Case:
Amazon Elastic Compute Cloud (EC2)

- Prominent IaaS provider
- Datacenters all over the world
- Many VM instance types
- Per-hour charging

<table>
<thead>
<tr>
<th>Instance</th>
<th>Capacity</th>
<th>US$/hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1.small</td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>m1.large</td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>cl.xlarge</td>
<td></td>
<td>0.76</td>
</tr>
</tbody>
</table>
Agenda

1. What is Cloud Computing?
2. IaaS Clouds, the Core Idea
3. The IaaS Owner Perspective
4. The IaaS User Perspective: How to Use Clouds? How to Choose Clouds?
5. Reality Check
6. Conclusion
Use Case: Workloads of Zynga  
(Massively Social Gaming)

Selling in-game virtual goods:

“Zynga made est. $270M in 2009 from.”
http://techcrunch.com/2010/05/03/zynga-revenue/

“Zynga made more than $600M in 2010 from selling in-game virtual goods.”
S. Greengard, CACM, Apr 2011

Sources: CNN, Zynga.
Source: InsideSocialGames.com
Use Case: Workloads of Zynga
(Massively Social Gaming)

• Load can grow very quickly
Provisioning and Allocation of Resources

Provisioning

Allocation

Load

Time

September 12, 2012
Provisioning and Allocation of Resources

Q: What is the interplay between provisioning and allocation?

Provisioning

Allocation

Load

Time

September 12, 2012
Provisioning and Allocation *Policies*

**Provisioning**
- When?
- From where?
- How many?
- Which type?
- etc.

**Allocation**
- When?
- Where?
- etc.

Q: How many policies exist?  Q: How to select a policy?

(Source: A. Antoniou, MSc Defense, TU Delft, 2012.)
Use Case:
Two Provisioning Policies, Compared

StartUp

OnDemand

Requested Instances
Accessible Instances

Use Case:
Two Provisioning Policies, Compared

Metrics for comparison

- **Job Slowdown ($JSD$):** Ratio of actual runtime in the cloud and the runtime in a dedicated non-virtualized environment

  $$Q: \text{Charged cost vs Total RunTime?}$$

- **Charged Cost ($C_c$)**

  $$C_c(W) = \sum_{i \in \text{leased VMs}} [t_{stop}(i) - t_{start}(i)]$$

- **Utility ($U$)**

  $$U(W) = \frac{SU_1(W)}{C_c(W)}$$

Use Case:
Two Provisioning Policies, Compared

Workloads

Uniform

Increasing

Bursty

## Use Case: Two Provisioning Policies, Compared

### Environments

<table>
<thead>
<tr>
<th>System</th>
<th>Hardware</th>
<th>VIM</th>
<th>Hypervisor</th>
<th>Max VMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAS4/Delft</td>
<td>20 Dual quad-core 2.4 GHz 24 GB RAM 2x1 TB storage</td>
<td>OpenNebula</td>
<td>KVM</td>
<td>64</td>
</tr>
<tr>
<td>FIU</td>
<td>7 Pentium 4 3.0 GHz 5 GB RAM 340 GB Storage</td>
<td>OpenNebula</td>
<td>Xen</td>
<td>7</td>
</tr>
<tr>
<td>Amazon EC2</td>
<td>unkown/Various</td>
<td>-</td>
<td>Xen</td>
<td>20</td>
</tr>
</tbody>
</table>

Use Case: Many Provisioning Policies, Compared

**Job Slowdown (J SD)**

Q: Why is OnDemand **worse** than Startup?

A: waiting for machines to boot

September 12, 2012
Use Case:
Many Provisioning Policies, Compared

Charged Cost ($C_C$)

Q: Why is OnDemand worse than Startup?
A: VM thrashing

Q: Why no OnDemand on Amazon EC2?

September 12, 2012
Use Case: Many Provisioning Policies, Compared

Utility ($U$)
Agenda

1. What is Cloud Computing?
2. IaaS Clouds, the Core Idea
3. The IaaS Owner Perspective
4. The IaaS User Perspective
5. **Reality Check:**  
   *Who Uses Public Commercial Clouds?*
6. Conclusion
The Real IaaS Cloud

VS

• “The path to abundance”
• On-demand capacity
• Cheap for short-term tasks
• Great for web apps (EI P, web crawl, DB ops, I/O)

• “The killer cyclone”
• Not so great performance for scientific applications (compute- or data-intensive)
Animoto: Video App on Amazon EC2

Scaled to peak of 3,500 instances in 3 days

Launch of Facebook modification

Number of EC2 Instances

Apr 12th  Apr 13th  Apr 14th  Apr 15th  Apr 16th  Apr 17th  Apr 18th  Apr 19th  Apr 20th
Zynga zCloud: Hybrid Self-Hosted/ EC2

• After Zynga had large scale
• More efficient self-hosted servers
  • Run at high utilization
• Use EC2 for unexpected demand

(Sources: http://seekingalpha.com/article/609141-how-amazon-s-aws-can-attract-ugly-economics and http://www.undertheradarblog.com/blog/3-reasons-zynga-is-moving-to-a-private-cloud/)
Other Cloud Customers

- 218 virtual CPUs
- 9TB/2TB block/S3 storage
- 6.5TB/2TB I/O per month

Customers in 190 Countries

(Source: http://markbuhagiar.com/technical/businessinthecloud/)
Agenda

1. What is Cloud Computing?
2. IaaS Clouds, the Core Idea
3. The IaaS Owner Perspective
4. The IaaS User Perspective
5. Reality Check
6. Conclusion
Conclusion: Take-Home Message

• **Cloud Computing** = IaaS + PaaS + SaaS

• **Core idea** = lease vs self-own
  - On-Demand, Pay-per-Use, Elastic, Pooled, Automated, QoS

• **The Owner Perspective**
  - Time-Sharing
  - Virtualization

• **The User Perspective**
  - Variable workloads
  - Provisioning and Allocation policies

• **Reality Check: 100s of users**

http://www.flickr.com/photos/dimitrisotiropoulos/4204766418/
Thank you for your attention!
Questions? Suggestions? Observations?

More Info:
- http://www.st.ewi.tudelft.nl/~iosup/research.html
- http://www.st.ewi.tudelft.nl/~iosup/research_cloud.html
- http://www.pds.ewi.tudelft.nl/

Alexandru Iosup
A.Iosup@tudelft.nl
http://www.pds.ewi.tudelft.nl/~iosup/ (or google “iosup”)
Parallel and Distributed Systems Group
Delft University of Technology

Do not hesitate to contact me…