LTAT.06.008
Cloud Computing
(6 ECTS)

Satish Srirama
satish.srirama@ut.ee
Course Purpose

• Introduce cloud computing concepts
• Introduce cloud services and providers
• Introduction to distributed computing algorithms like MapReduce
• Recent developments in the domain such as Serverless computing, fog computing etc.
• Future research directions – Cloud Manifesto

• https://courses.cs.ut.ee/2020/cloud/spring/
Schedule

• Lecture
  – Tuesday 14.15 - 16.00 (Narva mnt 18 - 1019)

• Labs
  – Tuesday 16:15 - 18:00 (Narva mnt 18 - 2010) (1. group)
  – Wednesday 16:15 - 18:00 (Narva mnt 18 - 2047) (2. group)
  – Wednesday 16:15 - 18:00 (Narva mnt 18 - 2048) (3. group)
Course Logistics

• Replacing Basics of Cloud Computing - MTAT.08.027 (Spring 2019) - 3 ECTS course
• Consult your curriculum coordinators
Questions

• Is everyone comfortable with data structures?
• How comfortable you are with algorithms?
• How comfortable you are with programming?
  – Java ?
    • External APIs?
  – Python – I assume you are
  – Web programming
Grading

• Written exam – 50%
• Labs – 45%
  – 14 lab exercises
• Active participation in the lectures (Max 5%)

• To pass the course
  – You need to score at least 50% in each of the above subsections
  – You need to score at least 50% in the total
Taking the Course

• People who have stuck with the curriculum have scored good

• Labs will get you easy points
  – Will try to provide bonus points for some of the labs

• Examination is considered to be a bit tough
  – One of the previous year’s examination paper (Basics of Cloud Computing - MTAT.08.027) is online

• Being a practical course it takes enough time and effort
Course schedule & Outline

• 11.02 Introduction to Cloud Computing
• 18.02 Platform as a service
• 25.02 IaaS and Virtualization
• 03.03 Serverless computing
• 10.03 Cloud storage services
• 17.03 Private and Hybrid clouds
• 24.03 Scaling enterprise applications on Cloud
• 31.03 Deployment models in cloud
• 07.04 Data processing on cloud
• 14.04 Information retrieval in cloud
• 21.04 Data warehousing
• 28.04 Data streams, data flow pipeline management
• 05.05 Fog computing
• 12.05 Cloud Manifesto
• 19.05 Examination 1 – Option 2
  – 18.05 Examination 1 – Option 1
Course schedule - continued

• Labs
  Labs follow the lectures
Grading of Labs

• Strongly suggest you to finish the lab exercises before coming to the next lecture
• Try to submit the results by Wednesday 16:00 next week to be graded for full score
• Late submissions
  – 10% will be deducted per day
  – You can submit exercises for 50% evaluation till end of the week after
Course schedule - continued

- **19.05** Examination 1 – Option 2
- **18.05** Examination 1 – Option 1

- Examination for second attempt – 2\(^{nd}\) June (Prospective)
Reference Books

• Mastering Cloud Computing: Foundations and Applications Programming
  Authors: Rajkumar Buyya, Christian Vecchiola, S.Thamarai Selvi

• Data-Intensive Text Processing with MapReduce
  Authors: Jimmy Lin and Chris Dyer

Reference Papers


• Relevant papers will be mentioned in respective lectures
"It’s nothing new"
"...we’ve redefined Cloud Computing to include everything that we already do..."

Larry Ellison, CEO, Oracle (Wall Street Journal, Sept. 26, 2008)

"It’s a trap"
"...marketing hype. Somebody is campaigning to make it true."


"It’s a trap!"
"...campaigning to make it true."

What is Cloud Computing?

• Computing as a utility
  – Utility services e.g. water, electricity, gas etc
  – Consumers pay based on their usage

1969 – Leonard Kleinrock, ARPANET project
  “As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of ‘computer utilities’, which, like present electric and telephone utilities, will service individual homes and offices across the country”
Timeline

Grid Computing
- Solving large problems with parallel computing
- Made mainstream by Globus Alliance

Utility Computing
- Offering computing resources as a metered service
- Introduced in late 1990s

Software as a Service
- Network-based subscriptions to applications
- Gained momentum in 2001

Cloud Computing
- Next-Generation Internet computing
- Next-Generation Data Centers
Clouds - Why Now (not then)?

• Experience with very large datacenters
  – Unprecedented economies of scale
  – Transfer of risk

• Technology factors
  – Pervasive broadband Internet
  – Maturity in Virtualization Technology

• Business factors
  – Minimal capital expenditure
  – Pay-as-you-go billing model
Virtualization

- Virtualization techniques are the basis of the cloud computing
- Virtualization technologies partition hardware and thus provide flexible and scalable computing platforms
- Virtual machine techniques
  - VMware and Xen
  - OpenNebula
  - Amazon EC2
Containers

• Use kernel of the host operating system instead of a hypervisor
  – Linux namespace
• Lightweight
  – No hypervisor overhead
  – Each Container acquires only required resources
• Fast start-up/ Bootup
  – Starting a container is faster than booting new OS or spinning up a new VM
• Performance
  – Near native performance
• Different container frameworks
  – LXC, Docker, Linux VServer, OpenVZ
  – Docker is an Open platform [https://www.docker.com/]
Cloud Computing - Characteristics

• Illusion of infinite resources
• No up-front cost
• Fine-grained billing (e.g. hourly)

• Gartner: “Cloud computing is a style of computing where massively scalable IT-related capabilities are provided ‘as a service’ across the Internet to multiple external customers”
Cloud Computing - Services

- **Software as a Service – SaaS**
  - A way to access applications hosted on the web through your web browser

- **Platform as a Service – PaaS**
  - Provides a computing platform and a solution stack (e.g. LAMP) as a service

- **Infrastructure as a Service – IaaS**
  - Use of commodity computers, distributed across Internet, to perform parallel processing, distributed storage, indexing and mining of data
  - Virtualization

- **SaaS**
  - Facebook, Flikr, Myspace.com, Google maps API, Gmail

- **PaaS**
  - Google App Engine, Force.com, Hadoop, Azure, Heroku, etc

- **IaaS**
  - Amazon EC2, Rackspace, GoGrid, SciCloud, etc.
Cloud Computing - Themes

- Massively scalable
- On-demand & dynamic
- Only use what you need - Elastic
  - No upfront commitments, use on short term basis
- Accessible via Internet, location independent
- Transparent
  - Complexity concealed from users, virtualized, abstracted
- Service oriented
  - Easy to use SLAs

SLA – Service Level Agreement
Cloud Models

- **Internal (private) cloud**
  - Cloud within an organization

- **Community cloud**
  - Cloud infrastructure jointly owned by several organizations

- **Public cloud**
  - Cloud infrastructure owned by an organization, provided to general public as service

- **Hybrid cloud**
  - Composition of two or more cloud models
Short Term Implications of Clouds

• Startups and prototyping
  – Minimize infrastructure risk
  – Lower cost of entry
• Batch jobs
• One-off tasks
  – Washington post, NY Times
• Cost associatively for scientific applications
• Research at scale
Economics of Cloud Users

- Many cloud applications have cyclical demand curves
  - Daily, weekly, monthly, ...
- Pay by use instead of provisioning for peak

![Diagram showing demand and capacity for static data center and data center in the cloud, highlighting unused resources.](image)
Economics of Cloud Users - continued

• Risk of over-provisioning: underutilization
  – Huge sunk cost in infrastructure

![Diagram showing static data center with unused resources](image)
Economics of Cloud Users - continued

- Heavy penalty for under-provisioning

  ![Diagram](image)
Economics of Cloud Providers

- Building a very large-scale datacenter is very expensive
  - $100+ Million (Minimum)
- Large Internet Companies Already Building Huge DCs
  - Google, Amazon, Microsoft...
- 5-7x economies of scale [Hamilton 2008]

<table>
<thead>
<tr>
<th>Resource</th>
<th>Cost in Medium DC</th>
<th>Cost in Very Large DC</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>$95 / Mbps / month</td>
<td>$13 / Mbps / month</td>
<td>7.3x</td>
</tr>
<tr>
<td>Storage</td>
<td>$2.20 / GB / month</td>
<td>$0.40 / GB / month</td>
<td>5.5x</td>
</tr>
<tr>
<td>Administration</td>
<td>≈140 servers/admin</td>
<td>&gt;1000 servers/admin</td>
<td>7.1x</td>
</tr>
</tbody>
</table>
Economics of Cloud Providers - continued

• **Power**

<table>
<thead>
<tr>
<th>Price per KWH</th>
<th>Where</th>
<th>Possible Reasons Why</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6¢</td>
<td>Idaho</td>
<td>Hydroelectric power; not sent long distance</td>
</tr>
<tr>
<td>10.0¢</td>
<td>California</td>
<td>Electricity transmitted long distance over the grid; limited transmission lines in Bay Area; no coal fired electricity allowed in California.</td>
</tr>
<tr>
<td>18.0¢</td>
<td>Hawaii</td>
<td>Must ship fuel to generate electricity</td>
</tr>
</tbody>
</table>

• **Cooling is also expensive**
  – Build data centers near rivers

• **Extra benefits**
  – Amazon: utilize off-peak capacity
  – Microsoft: sell .NET tools
  – Google: reuse existing infrastructure
Economics of Cloud Providers - Failures

• Cloud Computing providers bring a shift from high reliability/availability servers to commodity servers
  – At least one failure per day in large datacenter
• Why?
  – Significant economic incentives – much lower per-server cost
• Caveat: User software has to adapt to failures
  – Very hard problem!
• Solution: Replicate data and computation
  – MapReduce & Distributed File System (Will discuss later in Lecture 9)
## Adoption Challenges

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>Multiple providers &amp; Use elasticity to prevent DDoS attacks</td>
</tr>
<tr>
<td>Data lock-in</td>
<td>Standardization</td>
</tr>
<tr>
<td>Data Confidentiality and Auditability</td>
<td>Encryption, VLANs, Firewalls; Geographical Data Storage</td>
</tr>
</tbody>
</table>
# Growth Challenges

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data transfer bottlenecks</td>
<td>FedEx-ing disks, Data Backup/Archival</td>
</tr>
<tr>
<td>Performance unpredictability</td>
<td>Improved VM support, flash memory, scheduling VMs</td>
</tr>
<tr>
<td>Scalable storage</td>
<td>Invent scalable store</td>
</tr>
<tr>
<td>Bugs in large distributed systems</td>
<td>Invent Debugger that relies on Distributed VMs</td>
</tr>
<tr>
<td>Scaling quickly</td>
<td>Invent Auto-Scaler; Snapshots for conservation</td>
</tr>
</tbody>
</table>
## Policy and Business Challenges

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Opportunity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reputation Fate Sharing</td>
<td>Offer reputation-guarding services like those for email</td>
</tr>
<tr>
<td>Software Licensing</td>
<td>Pay-for-use licenses; Bulk use sales</td>
</tr>
</tbody>
</table>
Long Term Implications of clouds

• Application software:
  – Cloud & client parts, disconnection tolerance

• Infrastructure software:
  – Resource accounting, VM awareness

• Hardware systems:
  – Containers, energy proportionality
Cloud Computing Progress

Web startups
CS researchers
Enterprise apps, eg email
Educators

scientific & high-performance computing (HPC)

Armando Fox, 2010
This week in Lab (Homework)

• Introduction to IaaS & Getting keys
  – Work with Virtual Machines

• Study the following paper
Next lecture

• Platform as a service
  – E.g. Google AppEngine
References

• Several of the slides are taken from Prof. Anthony D. Joseph’s lecture at RWTH Aachen (March 2010)

• Papers to read