From Cluster to Cloud to Appliance

David J. Malan
malan@harvard.edu
CS50
Introduction to Computer Science I
prior cs courses

- three+: 2%
- two: 4%
- one: 19%
- zero: 75%
comfort level

- 14% less comfortable
- 13% somewhere in between
- 12% more comfortable

Year 2008:
- 34% less comfortable
- 43% somewhere in between
- 25% more comfortable

Year 2009:
- 44% less comfortable
- 42% somewhere in between
- 14% more comfortable

Year 2010:
- 46% less comfortable
- 46% somewhere in between
- 8% more comfortable

Year 2011:
- 55% less comfortable
- 35% somewhere in between
- 10% more comfortable

Year 2012:
- 60% less comfortable
- 30% somewhere in between
- 10% more comfortable
prior cs courses

- Zero: 56%
- One: 21%
- Two: 9%
- Three+: 14%
on-campus cluster
1990s – 2007
upsides of cluster

+ managed by university
+ familiar to students
downsides of cluster

- no root access
- outdated software
- no support after hours
off-campus cloud
2008 – 2010
cost $15 per student

<table>
<thead>
<tr>
<th></th>
<th>CPU</th>
<th>Disk</th>
<th>I/O Requests</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep</td>
<td>2,275 Hrs</td>
<td>125 GB</td>
<td>45,348</td>
<td>14 GB</td>
</tr>
<tr>
<td>Oct</td>
<td>3,425 Hrs</td>
<td>108 GB</td>
<td>93,257,314</td>
<td>191 GB</td>
</tr>
<tr>
<td>Nov</td>
<td>5,484 Hrs</td>
<td>199 GB</td>
<td>337,019,916</td>
<td>239 GB</td>
</tr>
<tr>
<td>Dec</td>
<td>5,206 Hrs</td>
<td>300 GB</td>
<td>427,639,962</td>
<td>52 GB</td>
</tr>
<tr>
<td>Jan</td>
<td>5,208 Hrs</td>
<td>300 GB</td>
<td>1,502,614,186</td>
<td>62 GB</td>
</tr>
</tbody>
</table>
upsides of cloud

+ root access
+ examine students' code in situ
+ no need for space, power, or cooling
+ more reliable
+ more scalable
+ topical for class
downsides of cloud

- time-consuming
- difficult (at the time) to configure
- unfamiliar to students
- latency
- mistakes made
client-side appliance

2011 –
```c
#include <stdio.h>

int main(void) {
    printf("hello, world");
}
```
implementation details

- Fedora Linux, Xfce
- Kickstart
- Boxgrinder
- RPMs (appliance50, cs51, ...)
- VirtualBox, VMware Fusion, VMware Player
- ...
- ...
performance

- fast
- fine (but computer fast)
- slow (but computer slow)
upsides of appliance

+ more familiar environment
+ graphical tools
+ distributed load
+ Apache, MySQL, ...
+ students could add software
+ anyone could engage actively
downsides of appliance

- time-consuming to prepare
- virtualization overhead
- brickable (hypervisor bugs)
key improvements

• TeamViewer
• Guest Additions, VMware Tools
• Dropbox
recommendations

• for large courses
• for departments
• for OpenCourseWare
• for web programming
cs50.net/appliance