Clouds on the Horizon: Evolving Teaching, Learning and Technology in Higher Education

Dr. Charles Severance
University of Michigan School of Information
October 6, 2009
Outline

• Looking at Google for inspiration
• Google App Engine
• Cloud Style High Performance Computing
• Cloud Intelligence
• Cloud Identity
• Cloud Learning
• A ToDo List of Tasks for IT and CS
Google
Capetown and George...

• Having an IMAP server in George is a pain

• GMail works no matter where he travels

• It is the *same* undersea cables, etc..

• Google is simply better than the Internet/Web
Google I/O 2008 Keynote

- Google I/O '08 Keynote by Marissa Mayer
- Usability / User Experience / User Testing / Architecture / Philosophy

http://www.youtube.com/watch?v=6x0cAzQ7PVs
Lessons

• The cloud is wide - we can touch 1000 servers in 0.1 seconds

• For things that seem “intelligent” 0.2 seconds is fast enough - as long as you can do a lot of them

• Lots of spread-out storage and a fast scan is important

• Data - Information - Knowledge - starts with data and the ability to look through that data quickly
Scalable Infrastructure

http://www.youtube.com/watch?v=zRwPSFpLX8I
• The only sustainable scalability is when you scale with inexpensive, green solutions
• Tape Backup is a rate limiting factor - so we need something creative
• Disaster recovery - “Of course!”
Google App Engine

- Expose Google’s worldwide Infrastructure to us as developers
- Make the web better
- Be the first widely used “cloud” environment - beat Amazon, Microsoft, and Yahoo!

http://www.youtube.com/watch?v=3Ztr-HhWXIc
http://www.youtube.com/watch?v=oTFL7FPLnXY
Google App Engine

- When you write a Google Application Engine Application - you are running in the Google Cloud
- Just like you were a Google Developer
- You don’t know where you are running or if one copy of a thousand copies of you are running
- Google hosts small applications for *free* - larger applications pay by usage
Hmmm...

- Clouds are an emerging new technical “form”
- Google is the first “prototype”
- Dive in...
- Teach a class
- Write a book
- Change the world?
Cloud-Style HPC
FIGURE 2. NPB-MPI (CLASS B) RUNTIMES ON 32 CPUS ON THE NCSA AND EC2 CLUSTER. BT AND SP WERE RUN WITH 16 CPUS ONLY. OVERLAID IS THE PERCENTAGE DEGRADATION IN THE EC2 RUNS.


90% utilization and low utility cost.

40% utilization and low utility cost.
What’s Faster? Amazon EC2 or a Supercomputer?

How would this graph change if we added the time it takes to apply for and be awarded resources on a supercomputer?


Note: Ian says that this is not precise research - it is to make us think.
The HPC Cycle...

- On campus scientists complain that it takes too long to get resources on national resources, or the queues are too long.
- We buy “strategic” on campus compute resources...
- We get as many scientists to try to use it as we can.
- Success means that local queues are really long and scientists complain that we need to get more processors.

The one constant in the cycle...
We do have a responsibility to help non-CS scientists make the best use of available HPC resources. Part of that training includes smaller local compute resources to be used to tune and adjust software....
Is the Singularity at Hand?

http://en.wikipedia.org/wiki/Technological_singularity
It is Getting Slightly Closer...

http://asimo.honda.com/
What is the Largest Number?

<table>
<thead>
<tr>
<th>25</th>
<th>1</th>
<th>114</th>
<th>117</th>
<th>150</th>
<th>152</th>
<th>120</th>
<th>46</th>
<th>19</th>
<th>126</th>
</tr>
</thead>
<tbody>
<tr>
<td>191</td>
<td>121</td>
<td>104</td>
<td>116</td>
<td>160</td>
<td>105</td>
<td>89</td>
<td>125</td>
<td>40</td>
<td>14</td>
</tr>
<tr>
<td>31</td>
<td>139</td>
<td>113</td>
<td>94</td>
<td>97</td>
<td>199</td>
<td>154</td>
<td>140</td>
<td>195</td>
<td>122</td>
</tr>
<tr>
<td>112</td>
<td>163</td>
<td>177</td>
<td>48</td>
<td>78</td>
<td>101</td>
<td>130</td>
<td>83</td>
<td>35</td>
<td>197</td>
</tr>
<tr>
<td>44</td>
<td>54</td>
<td>106</td>
<td>143</td>
<td>59</td>
<td>38</td>
<td>3</td>
<td>41</td>
<td>93</td>
<td>81</td>
</tr>
<tr>
<td>20</td>
<td>164</td>
<td>4</td>
<td>11</td>
<td>131</td>
<td>0</td>
<td>107</td>
<td>71</td>
<td>159</td>
<td>69</td>
</tr>
<tr>
<td>0</td>
<td>187</td>
<td>169</td>
<td>99</td>
<td>15</td>
<td>8</td>
<td>26</td>
<td>8</td>
<td>192</td>
<td>11</td>
</tr>
<tr>
<td>129</td>
<td>73</td>
<td>45</td>
<td>9</td>
<td>24</td>
<td>188</td>
<td>42</td>
<td>151</td>
<td>51</td>
<td>183</td>
</tr>
<tr>
<td>179</td>
<td>79</td>
<td>50</td>
<td>76</td>
<td>34</td>
<td>33</td>
<td>185</td>
<td>102</td>
<td>193</td>
<td>184</td>
</tr>
</tbody>
</table>
What is the Second Largest Number?
What are the last words of “Where the Wild Things Are”?
What are the last words of Where the Wild Things Are?

He then travels to where the wild things are, where they make him their king. .... Last words, (Click to show. Warning: May contain spoilers. ...)
And the point is...

- Researchers will increasingly want to write applications that seem intelligent.
- Each of those applications need a few terabytes of spinning, online permanent, network accessible, backed up data that I accumulate over time.
- The data will be images, sound, video, or text.
- I want to be able to scan all my data, asking a question and getting an answer in less than a second.
Your Assignment: Watch the movie “Eagle Eye” and decide what fraction of the technology in the movie will be commonplace within 20 years.
What Is Hadoop?

The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing. Hadoop includes these subprojects:

- **Hadoop Common**: The common utilities that support the other Hadoop subprojects.
- **Avro**: A data serialization system that provides dynamic integration with scripting languages.
- **Chukwa**: A data collection system for managing large distributed systems.
- **HBase**: A scalable, distributed database that supports structured data storage for large tables.
- **HDFS**: A distributed file system that provides high throughput access to application data.
- **Hive**: A data warehouse infrastructure that provides data summarization and ad hoc querying.
- **MapReduce**: A software framework for distributed processing of large data sets on compute clusters.
- **Pig**: A high-level data-flow language and execution framework for parallel computation.
- **ZooKeeper**: A high-performance coordination service for distributed applications.

http://hadoop.apache.org/

http://developer.yahoo.net/blogs/hadoop/
Cloud Identity

It is actually not that tough...
Unified Identity?

• The commercial cloud space seems to have given up on people having exactly one overarching identity - used everywhere.

• People do not *want* one identity but they also do not want to be forced to get one more identity.

• Use whatever you have and want to use - Google, Yahoo!, OpenId, Twitter, FaceBook, Blogger....

• Identity is more than just at the moment of login. OAuth.net
How to you handle WiFi guests on your campus?
Worldwide* Academic WiFi Exchange

www.eduroam.org

* Offer not valid in the United States
Eduroam Architecture

- **RADIUS** - Remote Authentication Dial-in User Service
- Created in 1991 by the Merit Network!
- RFC2866

http://en.wikipedia.org/wiki/RADIUS
Cloud Learning
EDU@2020

• Ricard N. Katz and Ronald Yanosky

• Google + Sakai = Google Phoenix

• Microsoft buys Blackboard

• Microsoft creates a lifelong personal portfolio

• Virtual teachers and virtual classmates

rtsp://educause.rmod.llnwd.net/a680/o1/edu2020.rm
Tsugi Project Goals

- Written in Python, Free, Open Source
- Hostable on Google for Free
- Easy to use learning tool building and hosting environment
- Supports IMS Learning Tools Interoperability
- Teachers writing tools - students writing tools (Thousands)

www.tsugiproject.org
wiscrowd.appspot.com

• A set of tools which allows multi-player educational in-class games

• Support for the teachers using the book “Wisdom of Crowds”

• Works either stand-alone or inside of an LMS using IMS Learning Tools Interoperability
### Available Applications

<table>
<thead>
<tr>
<th>Application</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Address Book</strong></td>
<td>Create a dynamic, online address book so your group can stay in touch. You control what fields are part of the address book.</td>
</tr>
<tr>
<td><strong>Pearson MyMathLab</strong></td>
<td>Provides practice problems sets and supporting instructional materials</td>
</tr>
<tr>
<td><strong>Favorites</strong></td>
<td>Enable your group to build a list of their favorite things. You can choose one of our default list types (favorite posts, movies, music, books, or TV) or create your own custom favorites type.</td>
</tr>
<tr>
<td><strong>For Sale</strong></td>
<td>Create an easy way for people to buy and sell from each other. You can choose to customize the application to make it about a specific type of item (e.g. bicycles or tickets or ) or leave it open to anything.</td>
</tr>
<tr>
<td><strong>Give Away</strong></td>
<td>Whether you're a Freecycling or Freesharing group that is dedicated to giving stuff away or you're a group that is about something else entirely but just wants to make exchanges possible among group members, this application makes it easier for folks to give stuff away.</td>
</tr>
</tbody>
</table>
If we assume tools can be created and hosted anywhere and Plugged into each other... What happens next?
Installing Your Next LMS

- Insert this code in your web page and press “Refresh”

```html
<script type="text/javascript" src="http://api.cloudsocial.org/js/ile-main.js"></script>
<script type="text/javascript">
ile_init("49c480149a008");
</script>
```
New: Learning Google Application Engine [www.appenginelearn.com](http://www.appenginelearn.com)

For the best effect to learn Python on your own, you should purchase the textbook and go through the materials in order. If you want to attempt the programming assignments make sure to install the appropriate software on your system. Installation instructions are provided under the "Software" tab.

This site should not be a substitute for a course you are taking – even if the course you are taking is using the same textbook. Each course and each instructor will take their own approach.

## Basic Python

- Writing Simple Programs (and Assignment Data)
- Computing with Strings (Chapter Through a file)
- Decision Structures (Chapter Through a File Again)
- Computers and Programs who has the most commits

---

http://www.cloudsocial.org/
Open Educational Resources

• Since CloudSocial allows a user to “bring their LMS along” when visiting content, it is an ideal tool to turn flat OER content into a collaborative learning space – without moving the content.

• Connextions – www.cnx.org

• Open Michigan – open.umich.edu

• MIT Open Courseware
If I worked in Campus IT...
"We don't just think outside the box... We transform the box into a robot and then think outside of the robot."
cPanel - Your Own PHP Cloud

• The software that powers ISPs like GoDaddy, RackSpace,

• Give it hardware - it includes and updates Linux, MySql, Admin, etc

• It deals with run-time performance and load balancing

• Export and import across cPanel-providers

• $20 per year per server for EDU - give every student an ISP
Outsourcing E-Mail: NO!

- I think everyone should have the choice to read their official campus E-Mail anywhere they like
- People need a low-level e-mail redirect including copy...
- Academic freedom / free speech
Disk Space - More

• Other than a network connection, my 10GB of AFS space at Michigan is my most valued IT benefit

• It is now full :( 

• I need about 3GB more per year - for media associated with my teaching output
A Few More Tasks...

• Make EduRoam happen here in the US - I would want to be the first school

• Lets clone Amazon EC2 and S3 protocols, etc and put up our own small clusters that talk these protocols - get involved in figuring out how to help scientists use these servers and protocols

• Start Looking at Hadoop - imagine a multi-institutional Hadoop instance - start with eliminating backup and giving us faculty more, cheap disk space - next move to data-intensive computing
Outline

- Looking at Google for inspiration
- Cloud Applications
- Cloud Computing
- Cloud Intelligence
- Cloud Identity
- Cloud Learning
- A ToDo List of Tasks for IT and CS