Exploratory Project: State of the Cloud, from University of Michigan and Beyond

Traci Ruthkoski
Office of Research Cyberinfrastructure
Goals

• Who is already in the cloud?
  – Excluding VMWare and HPC Services
  • Why?
    – Agility
    – Automation
    – Scalability

• Why not the cloud?
Results

Inter-University Consortium for Political and Social Research (ICPSR)

Cloud: Amazon EC2

Usage: Social Data Management and Presentation

Commonalities:
• Lack of local resources
• Usage model works well with cloud architecture

College of Engineering / School of Information

Cloud: Yahoo M45

Usage: Hadoop Research
Barriers to Entry

• Security & Regulation
  – Firewalls
  – Encryption
  – FDA, HIPAA, SAS70,…

• Data Transfer
  – Connectivity
  – Large data sets
  – Frequency

• Reliability
  – Uptime
  – Disaster Recovery

• Performance
  – Architecture

• Procurement
Unrealized Benefits

- Scientists, not Systems Administrators
- Efficient tools for IT organizations
Recommendations

• Categorization
• Security and Regulation
• Data Transfer
• Procurement
• Long-Term Development
Recommendations: Categorization

- **TIER-2: Mission Critical**
  - University & Department Operations / Research Activities
  - High Security Requirements
  - Local (Physical or VM)
    - Student Records
    - Faculty / Staff Records
    - Patient Records
    - Finance & Purchasing

- **TIER-1: Non-Mission Critical**
  - Research Activities
  - Medium Security Requirements
  - Low Security Requirements
  - Cloud / Grid
    - HPC
    - Data Storage
    - Desktop
    - SaaS (Matlab, SAS, Excel)
    - Data Storage (incl. media)
    - Web Services
    - CRM
    - Email
**Recommendations:** Categorization

### (Current) General Requirements
- Smaller data sets (< 5TB approx.)
- Data is not highly regulated by government policy
- Workflow does not require data transfer in/out of cloud storage on a regular basis
- Primary data access can be accomplished through cloud-hosted applications

<table>
<thead>
<tr>
<th>IaaS</th>
<th>PaaS</th>
<th>SaaS</th>
</tr>
</thead>
</table>
| **User / System Administration**
  User / System Administration requires access to operation system level configuration.  
  User / System Administration does not have local datacenter space to host servers.  
  Server requirements for performance can be met by cloud providers. | **User does not require system administrator access, but only requires a development platform.**  
  Development platform requirements can be fulfilled by cloud provided platform services. | **User is not interested or does not need to develop or administer software.**  
  Software required by user can be provided by cloud vendor software services. |

---

**MATLAB**

*IDE for applying custom code or SQL, .NET or Java or Rails*
## Recommendations: Categorization

<table>
<thead>
<tr>
<th>SaaS</th>
<th>PaaS</th>
<th>IaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>TeraGrid</td>
<td>VMware</td>
</tr>
<tr>
<td>MATLAB</td>
<td>DiaGrid</td>
<td>redhat.</td>
</tr>
<tr>
<td>Excel</td>
<td>Windows Azure</td>
<td>Network Satellite</td>
</tr>
<tr>
<td>SAS</td>
<td>SQL / .NET</td>
<td>Amazon Web Services</td>
</tr>
<tr>
<td>SPSS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recommendations: Security & Regulation

- Policy and SLA
- Policy
- SLA
- Policy and SLA
  - Policy and SLA
    - Policy and SLA
      » Policy and SLA
Recommendations: Data Transfer

• Any volunteers?

• Current pioneers:
  – Globus Online
  – Internet 2
  – SDSC
Recommendations: Procurement

Users Interact Directly With Cloud Provider

IT Organization As Service Broker

Campus IT Organization

Research Investigators

Research Investigators
Recommendations: Long-Term Development

• Partnerships
  – Academic with Academic
  – Academic with Industry

• Software strategy instead of hardware strategy
  – Holistic and appears simple to users
Conclusion

• The cloud is a solution, not the only solution.
• User need must fit architecture.
• Technical and social issues