IQSS Technical Advisory Committee
July 26, 2010
Social science research computing in the cloud
Cloud principles

- The cloud is a set of resources managed remotely by a third party.
- Cloud providers support administrator ratios as high as 1:1000 systems.
- There are several types of cloud services
  - Software as a Service (SaaS)
    - A vendor provides software for a specific function and the vendor stores data such that no data needs to be stored locally.
  - Platform as a Service (PaaS)
    - A vendor provides a framework to develop new applications and part of the framework involves storing state or performing certain functions for your application.
  - Infrastructure as a Service (IaaS)
    - A vendor provides basic compute, storage, and network infrastructure, leaving it up to the customer to manage those resources as if they were local resources.
IQSS technical services

- Dataverse: archive of research data sets
- OpenScholar: templates and hosting of personal / project web sites
- Research Computing Environment (RCE): cluster computing for statistical research
- Data hosting
- Network storage
- Web site hosting
- Desktop support
- Computer lab services
RCE in detail

• RCE = Research Computing Environment, remote infrastructure and scalable computing power for research analysis

• Used to run common statistical applications
  – R, GAUSS, Mathematica, MATLAB, Octave, SAS, S-PLUS, Stata

• RCE has three types of nodes
  – Login nodes
    • User logs in via NX (similar to VNC) and gets a desktop session
    • User can launch an application directly from the desktop
  – Compute-on-demand nodes
    • 25 compute cores on 9 nodes
    • User has special “RCE Powered Applications” menu to launch applications on machines with large memory resources (up to 64GB)
  – Batch nodes
    • 572 compute cores on 79 nodes
    • Used typically for non-interactive, long-running, scalable jobs
    • Most jobs use R
RCE architecture and configuration

Key Applications
- R
- GAUSS
- Mathematica
- MATLAB
- Octave
- SAS (COD only)
- S-PLUS
- Stata (SE and MP)

Interactive nodes

Login nodes

Compute-on-demand nodes

Batch nodes

Interactive nodes

Login nodes

Compute-on-demand nodes

Batch nodes

User

Secure login via NX

Request remote job

Allocate and manage resources

Local

Remote

Resource Manager

Key Applications
- R
- GAUSS
- Mathematica
- MATLAB
- Octave
- SAS (COD only)
- S-PLUS
- Stata (SE and MP)
Batch nodes and R
The case for RCE and the cloud

• To advance research computing technologies, we need to focus less on commodity services.
• External vendors manage large commodity clusters more efficiently than any in-house operation.
• Embarassingly parallel queries, the bulk of social science data analysis, are ideal research to benefit from the cloud resources.
• The RCE’s structure allows a gradual transition and hybrid infrastructure.
• Clouds will expand the range of hardware and platform support for all researchers.
RCE batch resource usage

![Graph showing resource usage over time.](image-url)
Cloud advantages

• Elasticity
  – Avoids over- and under-resourcing
  – Eliminates need to pay for resources not in use
  – Accesses much larger set of resources when needed

• Increased research computing focus
  – Offloads hardware maintenance to the “experts”
  – Focuses local staff on working with researchers to develop the next generation of social science computing tools

• Customized user environments
  – Sets up each cloud OS image with only the software needed

• More direct accounting of usage
  – Reduces divisional upfront commitment
  – Charges project for specific time / resources used
Cloud challenges

• Security
  – Restricting / controlling access to RCE cloud nodes
  – Dealing with confidential data

• Data hosting
  – Proximity to executing hosts
  – Latency and bandwidth
  – Backups
  – Availability

• Licensing
  – Variable underlying hardware

• Accounting
  – Setting individual and global limits
  – Monitoring limits
The hybrid model: RCE and the cloud

Key Applications

- R
- GAUSS
- Mathematica
- MATLAB
- Octave
- SAS (COD only)
- S-PLUS
- Stata (SE and MP)

- R
- GAUSS
- Mathematica
- MATLAB
- Octave
- SAS
- S-PLUS
- Stata (SE and MP)

- R
- Octave
Moving exclusively to the cloud

Challenges

1. Managing number of nodes in cluster
2. Securing communication between local and remote resources
3. Syncing local and remote data
4. Managing cost for high-memory nodes
5. License management and connection issues for interactive apps
Cloud vendor options

- **Microsoft Azure**
  - Current RCE user base and apps are Linux-based
  - More of a platform for building Windows applications

- **Google Apps**
  - Currently intended for specific apps (mail, calendar, etc.)

- **Amazon**
  - Very similar to current work environment for RCE
  - Uses Amazon Machine Image (AMI), essentially an OS image
  - Leverages existing vendors (Red Hat, VMware)

- **Other similar IaaS infrastructure offerings:**
  - Rightscale
  - Rackspace Cloud
  - ElasticHosts
Questions for the TAC

• Compute
  – How can we make the cloud easier for the users?
  – Are interactive sessions from the cloud realistic?
  – Would users tolerate slower single-node speed in exchange for increased scalability?
  – Should we focus more on ease-of-use or performance optimization?

• Data hosting
  – What would the tolerance be for data transfer time before execution?
  – Should we keep a local copy of data (live and/or backup)?

• Security
  – Can data access and transmission be comfortably isolated?
  – How should cloud compute resources be restricted in their use?
  – What other threats are users concerned about on the cloud?

• Accounting
  – What are the advantages / disadvantages of different payment models (pay-as-you-go, free services, etc.)?
RCE cloud roadmap

Batch Alpha
• prototype
• open source apps only
• manual data transfer

Batch Beta
• available on production cluster
• open source apps only
• manual data transfer
• sketch out future data model
• sketch out accounting structure
• force cloud use via special command
• no confidential or sensitive data

Interactive Alpha
• first “cloud-licensed” apps
• automated elasticity
• self-service mode
  – customized OS images
  – Data model (Phase I)
  – User accounting (Phase I)

Interactive Beta
• provide as open source
• self-service mode
  – Data model (Phase II)
  – User accounting (Phase II)