StratusLab

C. Loomis (CNRS/LAL)

EGEE'09 (Barcelona)

21 September 2009

www.eu-egee.org

Information Society and Media

StratusLab

- **Fusion of grid and cloud technologies**
  - Grid: Federation of resources, collaborative services
  - Cloud: Elasticity, customized environments for users

- ***Open*** **collaboration between 6 partners**
  - http://stratuslab.org/

Amazon Web Services

- **Elastic Compute Cloud (EC2)**
  - Provides easy interface for allocating and starting machines.
  - Five different machine configurations available.
  - Supports various *Linux*, OpenSolaris, Windows server.

- **Elastic Block Store (EBS)**
  - Acts as a mountable, persistent disk.
  - Can associate disk with image dynamically at startup.
  - Uses Simple Storage Service (S3) behind the scenes.

- **Elastic IP**
  - Persistent allocation of IPv4 address.
  - Can associate image instance and address dynamically.

Fabric Management Tools

- **For wide-spread adoption, cloud resources must be compatible with standard resource center fabric management tools.**

- **Major gLite services all deployed in AWS with Quattor**
  - http://quattor.org/
  - Challenges related to late knowledge of execution "context" parameters (network, hostname, …)
  - Minor modification for fabric "installation" phase
  - Offers easier mechanism for deploying large number of "identical" machines (e.g. worker nodes)

Expanding a Grid Site

- **Cloud resources can be used to dynamically add new resources to existing sites.**

- **Experiment was done by running additional worker nodes for a production EGEE site in AWS.**
  - Additional workers used for several months in a production site
  - No real issues encountered either from user or system administrator points of view

Grid Services

- **Deploying all grid services in the cloud is a step to running resource centers on "private" clouds.**

- **BDII (information system)**
  - Usual virtualization/Xen issues (e.g. TLS libraries)
  - Problem with self-configuration of hostname because of "bizarre" network configuration of machines

- **Secured Services**
  - All other services deployed but not tested
  - Likely issue with reverse-DNS lookups in GSI because these cannot be altered in commercial clouds

Issues

- **Contextualization**
  - Late availability of configuration information for machines
  - Standard needed for finding/using dynamic information
- **Support of virtualization**
  - Routine testing of services on virtualized resources
  - Treat these resources as standard supported platforms
- **Reverse DNS**
  - Commercial clouds will (likely) not provide this.
  - Modify GSI settings globally to work around issue?
  - Mark virtualized resources to allow users to configure/choose?

Conclusions

- **Cloud technologies offer attractive mechanism for providing underlying infrastructure.**
  - No real technical show stoppers. Reverse-DNS could pose a policy problem.
  - Need active testing of services on virtualized resources.
- **Standardization would help the adoption of cloud technologies within the grid.**
  - Common set of APIs (although AWS is a *de facto* standard)
  - Standard mechanism for providing machine context
- **Look also at other abstractions for providing services.**
  - E.g. Hadoop is now available in AWS.
  - Could replace pilot job/task management frameworks.