Federating Infrastructure as a Service cloud computing systems to create a uniform e-infrastructure for research

David C.H. Wallom¹, Matteo Turilli¹,², Michel Drescher³, Diego Scardaci³,⁴ and Steven Newhouse³,⁵

¹University of Oxford, Oxford, UK
²Rutgers University, New Jersey, USA
³EGI.eu, Amsterdam, NL
⁴INFN Division of Catania, Catania, Italy
⁵EMBL-EBI, Hinxton, UK
Overview

• Why?
• What?
• Who?
• Future?
Success with the Grid
Rationale

Growth of Providers
• Grid Computing
  – Academic resource providers
• Cloud Computing
  – Diversity of resource providers

Growth of Research Communities

Tens of 1000’s
Few related use cases
Single application model

Millions
Many diverse use cases & application models
Principles of Federation

• **Provider agnosticism**: the only condition to federate resources is to expose the chosen interfaces and services.

• **Heterogeneous implementation**: no mandate on the cloud technology, protecting resource provider investment

• **Standards and validation**: Use only recommended and common open standards for the interfaces and images.

• **Resource integration**: Cloud Computing to be integrated into the existing production infrastructure, protecting prior EGI investment and increasing ROI
User-offer Target

- **Total control** over deployed applications
- **Elastic resource consumption** based on real needs
- **Workloads** processed on-demand
- **Endorsed and accredited applications** available from multiple different communities shared
- **Single sign-on** at multiple, independent providers
- **Centralised access** to service information across multiple providers
<table>
<thead>
<tr>
<th>#</th>
<th>Capability</th>
<th>User Stories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Virtual Machine management</td>
<td>“I want to instantiate a single existing VM image on a remote cloud.”</td>
</tr>
</tbody>
</table>
| 2  | Data management                         | “I want to instantiate a VM instance from an image that I have created and is not on the cloud I wish to use.”  
                              | “I want to associate my running VM with a data set in the Cloud.”  
                              | “I want to take snapshots of my running VM for restart purposes.”                                                                         |
| 3  | Integrated information system           | “I want to choose on which resource provider I want to start my single VM.”  
                              | “I need to know about the Virtual Machine Manager (VMM) capabilities the provider offers.”                                                |
| 4  | Accounting                              | “My usage across different resource providers needs to be recorded and reported to multiple aggregators.”                                    |
| 5  | Availability & Reliability              | “Information relating to the availability/reliability and current status of the remote virtualised resource needs to be available to me.” |
| 6  | VM & Resource state change notification | “When the status of the [VM] instance I am running changes (or will change) I want to be told about it.”                                    |
| 7  | Integrated AAI                          | “I want to use my existing identity, and not re-apply for new credentials to use each component of the service.”                              |
| 8  | VM Image Management                     | “I want to use a single VM image across multiple different infrastructure providers.”                                                          |
| 9  | Brokering                               | “I want my VM instance to run on a resource that is suitable based on a set of policies or requirements rather than my choosing directly which resource will run it.” |
| 10 | Contextualisation                       | “When I deploy a VM instance on a resource I want to give it configuration information for customisation of the default template. This can only happen when it is up and running.” |
Cloud within the EGI Infrastructure Platforms

User Community

EGI Cloud Infrastructure Platform

Uniform interfaces to Cloud Compute and Storage

Cloudinit
OCII
CDMI

EGI Cloud Infrastructure Platform

Cloud Management Stacks
(OpenStack, OpenNebula, Synnefo, ...)

Monitoring and control of utilisation

Technical Consultancy and Support

EGI Core Platform

Secure endorsement, Application and Service Delivery

EGI Collaboration Tools

OVF

Secure endorsement Application and Service Delivery

EGI Collaboration Tools

Cloudinit
CDMI

Monitoring and control of utilisation

Technical Consultancy and Support

EGI Core Platform

Secure endorsement, Application and Service Delivery

EGI Collaboration Tools

OVF

Secure endorsement Application and Service Delivery

EGI Collaboration Tools

Cloudinit
CDMI

Monitoring and control of utilisation

Technical Consultancy and Support

EGI Core Platform

Secure endorsement, Application and Service Delivery

EGI Collaboration Tools

OVF

Secure endorsement Application and Service Delivery

EGI Collaboration Tools

Cloudinit
CDMI

Monitoring and control of utilisation

Technical Consultancy and Support

EGI Core Platform

Secure endorsement, Application and Service Delivery

EGI Collaboration Tools

OVF

Secure endorsement Application and Service Delivery

EGI Collaboration Tools

Cloudinit
CDMI

Monitoring and control of utilisation

Technical Consultancy and Support

EGI Core Platform
## Cloud technology Integration Level

<table>
<thead>
<tr>
<th>Cloud Management Stack</th>
<th>Fed. AAI</th>
<th>Monitoring</th>
<th>Accounting</th>
<th>Img. Mgmt.</th>
<th>OCCI</th>
<th>CDMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenStack</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>OpenNebula</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Synnefo</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>WNoDeS</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>StratusLab*</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloudstack</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Emotive</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>
Resource Allocation

- e-Grant supports;
  - Resource request
  - Resource availability
- 5,000 job slots and 170 TB of storage aside

- Preliminary Units of Allocation

<table>
<thead>
<tr>
<th>Name</th>
<th># of vCPU</th>
<th>Mem (GB)</th>
<th>Storage (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>1</td>
<td>2</td>
<td>1 x 20</td>
</tr>
<tr>
<td>Medium</td>
<td>2</td>
<td>4</td>
<td>1 x 40</td>
</tr>
<tr>
<td>Large</td>
<td>4</td>
<td>8</td>
<td>2 x 80</td>
</tr>
<tr>
<td>Other</td>
<td>&gt;2</td>
<td>&gt;7.5</td>
<td>n x &gt;40</td>
</tr>
</tbody>
</table>
FedCloud Infrastructure

• One year of production

• Resources
  – 21 providers from 14 NGIs (Countries)
    • 55% Openstack, 42% Open Nebula, 3% Syneffo
  – 17 interested in joining from 7 new NGIs

• Usage
  ~700K VMs
  ~18M CPU hours
Supported Usage Models

• **Service Hosting**
  – Long-running services (e.g. web, database or application servers)

• **Compute and data intensive workloads**
  – Batch and interactive (e.g. IPython, R, MATLAB) with scalable and customized environments not limited to the traditional job model

• **Datasets repository**
  – Store and manage large datasets for your applications

• **Disposable and testing environments**
  – Host training events, test new developments and applications without overhead
## Support Model

### Status Description

<table>
<thead>
<tr>
<th>Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-assessment</td>
<td>FCUS members identified a potential use case that can profit from the EGI Federated Cloud services</td>
</tr>
<tr>
<td>Assessed</td>
<td>The use case requirements are assessed with relevant information added to a dedicated wiki page, to manage the full porting &amp; integration process</td>
</tr>
<tr>
<td>Preparatory</td>
<td>The user &amp; FCUS team setup the environment to execute tests on the FedCloud (configuration of client environment, identify resource providers, upload vm images, …)</td>
</tr>
<tr>
<td>Test &amp; Integration</td>
<td>Users are accessing the FedCloud through the fedcloud.egi.eu VO to execute tests while integrate their applications to the infrastructure</td>
</tr>
<tr>
<td>Pre-production</td>
<td>Test &amp; integration phases were successfully completed. Create a production Virtual Organization (VO) or join an already existing VO</td>
</tr>
<tr>
<td>Production</td>
<td>Use case is being used in production. The use case completed all tests and is regularly making use of the Federated Cloud using a production level VO</td>
</tr>
<tr>
<td>Closed</td>
<td>Test &amp; integration successfully completed. Use case does not foresee moving into production.</td>
</tr>
<tr>
<td>Cancelled</td>
<td>Test &amp; integration did not successfully completed. User cancelled the use case.</td>
</tr>
</tbody>
</table>
Use Case Discipline Classification

Usecases
- 12 @ Launch
- 60 to date
  - 15 production
**READemption**

- **Scientific Discipline:** Natural Science, Biological Sciences, Bioinformatics
- **Status:** Production (hightroughputseq.egi.eu VO)
- **Sites:** GoeGrid, IFCA (Nov-Dec 2014)

**Next Generation Sequencing**
- Implemented in Python 3
- Needs:
  - Third-Party libraries (numpy, scipy, matplotlib, pysam)
  - Short Read mapper ‘Segemehl’
  - R
- Usual runtime on a multi-core machine several hours to some days

**Implementation on EGI Federated Cloud**
- VO supported by GWDG, IFCA and CESNET sites
- VM with 24 cores and 128 GB of RAM
- Block storage up to 3TB
- To serve peak loads

Source: Konrad U. Förstner
The Ecological Niche Modeling (ENM) Workflow takes as input a file containing species occurrence points to create a model with the openModeller Web Service.

- The EUBrazilOpenBio ENM service is exposed through an extended openModeller Web Service interface
- Multi-staging and multi-parametric oM experiments are implemented through COMPSs that dynamically creates the virtual resources to execute the operations.
- An OCCI connector is used for the VMs management while data management supports CDMI endpoints.

Service available at https://portal.biovel.eu/
Peachnote is a **music score search engine and analysis platform**.

Hundreds of thousands of music scores are being digitized by libraries all over the world. In contrast to books, they generally remain inaccessible for content-based retrieval and algorithmic analysis. There is no **analogue to Google Books for music scores**, and no large corpora exists that can empower advanced analysis on music scores. Peachnote **want to help change that** providing visitors and researchers access to a massive amount of symbolic music data.
Chipster

- **Scientific Discipline**: Natural Science, Biological Sciences, Bioinformatics
- **Status**: Test & Integration (fedcloud.egi.eu VO)
- **Sites**: PRISMA-INFN-Bari

**ELIXIR Pilot Action Proposal:**
Using virtual machines and clouds in bioinformatics training

User-friendly analysis software for high-throughput data:
- NGS
- Microarray
- Proteomics
- sequence data

**Chipster in the EGI FedCloud:**
- ‘light’ VM (datasets removed)
- Chipster VM configured through contextualisation
- shared block storage exported as NFS for tools (500 GB)
- block storage for output (500 GB)
• Inter and Intra cloud networking,
  – Using standards to support creation of custom multi resource private networks for cloud resources

• Development of multi provider and multi consumer business models for federated cloud operation,

• Expansion in higher level tools and services building on concrete foundations.

• Further development and release of updated standards in cloud and other related areas
  – In development progress of complete EGI profile which will fully document all interfaces to the CMF

• Connect resources from other commercial cloud providers to showcase deployment of standards
Federated Cloud Services

Tier 1:
Reliable Infrastructure Cloud

Federated IaaS and STaaS Cloud

Tier 2:
General-purpose platform services

Tier 3:
Platform as a Service

Tier 4:
Zero ICT Infrastructures
Conclusion

The EGI Federated Cloud is a federation of institutional private Clouds, offering Cloud Services to researchers in Europe and worldwide.

A single cloud system able to
• Scale to user needs – beyond test and development instantiations to production private clouds
• Integrate multiple different providers to give resilience
• Prevent vendor lock-in
• Enable resource provision targeted towards the research community
Thanks!

Thanks & Questions?