Compute Services in EGI
Overview and use cases

Enol Fernández
enol.Fernandez@egi.eu

The work of the EGI Foundation is partly funded by the European Commission under H2020 Framework Programme
Session outline

- **EGI Compute Services**
  - Cloud Compute
  - Cloud Container Compute
  - High Throughput Computing

- **Users experiences**
  - NextGEOSS EO data processing campaigns on EGI.eu Federated Cloud
  - AiiDA lab - an ecosystem for developing, executing, and sharing scientific workflows
  - Simulation of high power laser experiments by using high performance computer

- **Q & A**
Run your workloads on the distributed resources on the EGI Federation
IaaS gives you API-based deployment of virtual machines (VMs) on-demand:

- control the hardware (CPU, memory, disk) and software (OS & any other software stacks) configuration
- without managing physical servers
Containers provide OS-level virtualisation

- Isolated runtime built on namespaces and cgroups, without VMM overhead
Docker commoditizes containers

- Hides and automates container management process
- One-command-line deployment of applications
- Easy to move from development to production
- Provides ecosystem to create and share images

“Open-platform for building, shipping and running distributed applications”
Kubernetes

Kubernetes is an open-source platform for automating deployment, scaling, and operations of application containers across clusters of hosts, providing container-centric infrastructure.

Kubernetes:
- Places containers on nodes
- Recovers automatically from failure
- Basic monitoring, logging, health checking
- Enables containers to find each other
Kubernetes

Container orchestration

- Schedule containers to physical or virtual machines
- Restart containers if they stop
- Provide private container network
- Scale up and down
- Service discovery
Jobs are a command (or series of commands) managed by a scheduler

- Unprivileged (no servers) and time limited (no long running services)
- Capable of delivering large amounts of processing capacity over long periods of time.
<table>
<thead>
<tr>
<th></th>
<th>Cloud Compute</th>
<th>Cloud Container Compute</th>
<th>High Throughput Compute</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is it?</strong></td>
<td>Distributed IaaS</td>
<td>Kubernetes on top of EGI Cloud Compute</td>
<td>The <em>grid</em>, a scalable batch system</td>
</tr>
<tr>
<td><strong>What you run?</strong></td>
<td>VMs</td>
<td>(Docker) Containers</td>
<td>Jobs</td>
</tr>
<tr>
<td><strong>Typical workloads</strong></td>
<td>Lift and shift existing applications, Specific OS (kernel) requirements, Long running servers</td>
<td>Cloud-native containerised applications.</td>
<td>Execution of parallel computing tasks to analyse large datasets.</td>
</tr>
<tr>
<td><strong>Pros / Cons</strong></td>
<td>[+]: Complete control over resources, run (almost) anything you’d like, [-]: Complex operation</td>
<td>[+]: Industry standard, [+]: Hides complexity of Kubernetes setup, [-]: Kubernetes steep learning curve, [-]: Complexity of Kubernetes setup</td>
<td>[+]: No management of resources, just submit jobs, [-]: Legacy interfaces, [-]: Jobs may not match any computational need</td>
</tr>
</tbody>
</table>

**Configurability**  
**Abstraction**
EGI Cloud Compute
Run virtual machines on-demand with complete control over computing resources
Distributed Infrastructure as a Service (IaaS) powered by the EGI Federated Cloud

- Allows international collaborations to perform distributed data analysis with VM-based workloads

Features:
- Execution of VMs on a distributed infrastructure
- Federated identity
- Common VM image catalogue
- GUI and CLI/API based access
- Support for IaaS orchestration
- Central accounting and monitoring
Access layers

GUI Access
- AppDB VMOps

Federated Access
- IaaS Orchestration
- Terraform

Direct API Access
- IaaS API
- IaaS API

AAI: Check-in

GUI Users
Developers/Advanced users
AppDB VMOps

A Federated GUI

Single Web dashboard to manage VMs in the federation

- Point-and-click wizard solution to create new VMs

Integrated with all the features of the federation

Powered by Infrastructure Manager orchestrator
IaaS Orchestration

Manage heterogeneous providers

Deploy IaaS resources on all kind of IaaS providers (including OpenStack)
TOSCA standard support
Integration with EGI Cloud features (image catalogue, authentication)

Open Source tool with support for multiple IaaS providers (including OpenStack)
Near API-level abstraction, but useful for interacting with different providers in a uniform way
IaaS access

Full access to OpenStack APIs with federated identity
Virtual Machines Images and AppDB

*Bring your software to the cloud*

**Common registry for Virtual Appliances (VA)**
- VM image + metadata
- Available for running at the EGI providers or on any hypervisor

**Community-level** management of VAs
- Get the needed software on all providers automatically
- Control what can be executed
Containers on EGI Cloud Compute

Run docker on VMs

Pre-built image with Docker, Docker-compose and Kubernetes installed
Instantiate at one provider, ssh and:

```
$ docker run ...
```
EGI Cloud Container Compute

Run docker containers on-demand with complete control over computing resources
Automated provision of Kubernetes clusters on EGI Cloud Compute providers

- Built on EC3 for scalable management of resources

1. Provision cluster VMs
2. Deploy Kubernetes cluster on provisioned VMs
3. Use native Kubernetes tooling
EC3: Elastic Cloud Computing Cluster

Deploys and configures Kubernetes

- Elasticity – can grow/shrink depending on your load
- GUI and CLI access
- Integrated with Check-in and AppDB

Builds on ansible and kubeadm
Container registry service for sharing applications

DockerHub is the main registry available

- Public and private repositories
- Official repositories
- Automated builds
- Free accounts conditions recently changed!
Helm is a package manager for Kubernetes

Allows to easily install applications/services easily

Creates from templates all the k8s objects needed to run the application

Wide catalogue of *charts* with (normally) sane defaults
EGI High Throughput Compute

Execute thousands of computational tasks to analyse large datasets
The EGI High-Throughput compute (HTC) provides users with the capability to access large amounts of computing resources, and to submit hundreds or thousands of computational tasks.
CernVM FileSystem provides scalable software distribution across the federation

- POSIX read-only filesystem in user space
- Uploader interface for software managers of each CVMFS repository
- Available at all providers of EGI HTC
  - Files cached at providers for fast access to frequently used software
Containers as jobs

Using HTC for running containers

❌ Docker
- Requires root access to run your container

✅ Singularity
- Container image and runtime designed to run on multi-tenant computational resources
- Conversion of images from docker (automatic)
- Requires system administrator to configure

✅ udocker
- run applications encapsulated in docker containers without using docker, without using privileges, without system administrators intervention and without additional system software
- Run as a normal user, with the normal process controls and accounting, in interactive or batch systems
The Workload Manager service is based on DIRAC technology and is suitable for users that need to exploit distributed resources in a transparent way.

- User-friendly interface and API access for advanced usage
- Combine Cloud / HTC providers for job execution
- Optimise workload placement across all kind of resources

Learn more tomorrow 4th Nov @ 13.15 on the **Data analytics services in EGI - Overview and use cases** session!
GPUs

Accelerators on EGI Compute services
Cloud Compute
• Specific VM *flavors* with GPUs available at CESNET-MCC, IFCA-LCG2, IISAS, and NCG-INGRID-PT providers, more coming in the next months
• VM images with GPU drivers ready to use

Cloud Container Compute
• Docker plugin and kubernetes configuration also ready

High Throughput Compute
• Selected sites support jobs with GPUs, just tag the jobs
EGI Compute Services

- Cloud Container Compute
- Cloud Compute
- Workload Manager
- HTC
- Container registry/helm
- AppDB
- Online Storage
- CVMFS
- EGI federation services
  - AAI (Check-in) / Accounting / Monitoring
## Summary

<table>
<thead>
<tr>
<th></th>
<th>Cloud Compute</th>
<th>Cloud Container Compute</th>
<th>High Throughput Compute</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is it?</strong></td>
<td>Distributed IaaS</td>
<td>Kubernetes on top of EGI Cloud Compute</td>
<td>The <em>grid</em>, a scalable batch system</td>
</tr>
<tr>
<td><strong>What do you run?</strong></td>
<td>VMs</td>
<td>(Docker) Containers</td>
<td>Jobs</td>
</tr>
<tr>
<td><strong>Typical workloads</strong></td>
<td>Lift and shift existing applications Specific OS (kernel) requirements Long running servers</td>
<td>Cloud-native containerised applications.</td>
<td>Execution of parallel computing tasks to analyse large datasets.</td>
</tr>
<tr>
<td><strong>Pros / Cons</strong></td>
<td>[+] Complete control on resources, run (almost) anything you’d like [-] Complex operation</td>
<td>[+] Industry standard [+Hides complexity of Kubernetes setup [-Kubernetes steep learning curve</td>
<td>[+ No management of resources, just submit jobs [- Legacy interfaces [- Jobs may not match any computational need</td>
</tr>
</tbody>
</table>

### Configurability vs. Abstraction
Thank you for your attention.

Questions?