The DODAS Experience on the EGI Federated Cloud

Daniele Spiga INFN (spiga@infn.it)
On behalf of Enol Fernàndez¹, Vincenzo Spinoso² & DODAS Team

ecs-hub.eu @EOSC_eu

24th International Conference on Computing in High Energy & Nuclear Physics, Adelaide (AU), 4-8 Nov 2019

¹ EGI.eu; ² INFN/EGI.eu
A quick introduction to
- EGI
- Dynamic On Demand Analysis Service (DODAS)

Outline

- Overview of EGI Resources exploitation using DODAS
  - The integration approach
  - Managing AuthN/Z
- Testing the system with CMS Experiment analysis workflows
  - Results
- Future and related work
EGI: Advanced Computing for research

Mission
Create and deliver open solutions for research and innovation by federating digital capabilities, resources and expertise between domains and across geographical/organisational boundaries.

Vision
Researchers and Innovators from all disciplines have easy, integrated and open access to the advanced scientific computing capabilities, resources and expertise needed to collaborate and to carry out data/compute intensive applications.
EGI is a federation of over 200 computing and data centres spread across Europe and the rest of the world.

- 47 Countries
- 71,000 users
- 12 Integrated e-Infrastructures
- 1,700 Open Access Publications in 2018
- 31 large-scale research collaborations
- 20+ business use cases
## Compute

- **Cloud Compute**
  - Run virtual machines on demand with complete control over computing resources

- **Cloud Container Compute BETA**
  - Run Docker containers in a lightweight virtualised environment

- **High-Throughput Compute**
  - Execute thousands of computational tasks to analyse large datasets

- **Workload Manager BETA**
  - Manage computing workloads in an efficient way

## Storage and Data

- **Online Storage**
  - Store, share and access your files and their metadata on a global scale

- **Archive Storage**
  - Back-up your data for the long term and future use in a secure environment

- **Data Transfer**
  - Transfer large sets of data from one place to another

## Security

- **Check-in BETA**
  - Login with your own credentials

## Applications

- **Applications on Demand BETA**
  - Use online applications for your data & compute intensive research

- **Notebooks BETA**
  - Create interactive documents with live code, visualisations and text

## Training

- **FitSM Training**
  - Learn how to manage IT services with a pragmatic and lightweight standard

- **ISO 27001 Training**
  - Learn how to manage and secure information assets

- **Training Infrastructure**
  - Dedicated computing and storage for training and education
Bringing new pilots to the EGI Cloud

Access to prototyping VOs

- Create Check-in account & enrol VOs
- Opportunistic non-guaranteed resources

Pilot development & evaluation

- Support from EGI support team and cloud providers
- Add any required base-images to VO

Production setup

- Create production VOs
- Negotiate SLA/OLAs with providers

03/11/2019

spiga@pg.infn.it

CHEP 2019, 4-8 November, Adelaide Australia, AU
**Dynamic On Demand Analysis Service: DODAS**

- A Platform as a Service whose aim is to guarantee deployment of complex and intricate setup on “any cloud provider” with almost zero effort.
- **Infrastructure as code**: driven by a templating engine to specify high-level requirements
- Allows to instantiate on-demand container-based clusters to execute software applications

DODAS is a Thematic Service under EOSC-hub H2020 EU project.

- **Resources abstraction**
- **Automation**
- **Multi Cloud support**
- **Federated Authentication**

- **Big Data Pre-Post processing**
  - HDFS, Spark
- **Exploitation of Machine Learning**
  - Training facility
  - Inference engine
- **Batch System as a Service**
  - HTCondor batch system
  - HTCondor federation solutions

---

**Evolution of an INDIGO-DataCloud use-case based on CMS workflow.**
A highly customizable system

DODAS is *experiment agnostic*
- Flexible enough to support multiple and diverse use cases

**Highly Customizable:** By design, DODAS provides **three major handles**

- **Docker**
  - To support user tailored computing environments

- **Ansible**
  - To automate configuration and deployment of custom services and/or dependencies

- **TOSCA**
  - To define input parameters and customize the workflow execution
  - To abstract from (hybrid) IaaS infrastructures
Exploiting EGI Resources through DODAS

- The underlying distributed compute infrastructure is provided by the EGI Federated Cloud: **5 providers selected**

- DODAS is used to
  - Describe the **infrastructure as a code**
  - Deploy the **virtual infrastructure** (details later)

---

**DODAS Deployer**

**EGI FedCloud**

**PaaS Federation Layer**

**IaaS Federation**

**TOSCA**

**Cloud Manager**

**Cloud Provider1**

**Cloud Provider2**

**HTCondor Overlay**

---

spiga@pg.infn.it

CHEP 2019, 4-8 November, Adelaide Australia, AU
The EGI FedCloud relies on the EGI Check-in service for authentication and authorization:
- Single Sign-On through eduGAIN, social media and other institutional or community-managed identity providers
- Harmonised authorisation information, aggregated from multiple sources, based on the VO concept
- Industry Standard OpenID Connect technology allowing web and non-web access to services

DODAS relies on a dedicated INDIGO-IAM service instance
- VO management, enrollment, authentication and authorization at DODAS core services
- EGI Check-in integrated as an external authentication mechanism

We use IAM as an AAI harmonization layer, so that:
- Check-in grants access to the EGI cloud resources
- The IAM harmonised identity is used to manage DODAS domain-specific security
  - e.g. to manage HTCondor security via GSI (see later)

See: Beyond X.509: token-based authN/Z in practice
Deploying a CMS virtual site with DODAS

- **HTCondor Glidein in the Vacuum**
  - Seamlessly integrating the CMS Submission Infrastructure
  - Include also CVMFS, WorkerNode, X509 micro service, Squids
- Virtual infrastructure managed by an elastic Kubernetes
- Automatically and dynamically created by DODAS

1. Gets IAM Token as incoming auth credential
2. Implements security via IAM token exchange
3. Cache and return X509 certificates to grant access to CMS
**Functional test: the CMS perspectives**

Integration with the Analytics and Monitoring Infrastructure at CERN (MONIT)

- The virtual infrastructure, implemented through DODAS, is seen as a single site by CMS.
- For the scope of this test we used **T3_IT_Opportunistic_dodas**

---

The DODAS virtual site wrt CMS Tier3

- **400 Running jobs**
- Sites joining the virtual infrastructure
- Colors are CMS Schedds the glideins connect to

---

**Data to clients**

**Data from remote**

- **Smart caching at CMS: applying AI to XCache edge services**

---

spiga@pg.infn.it

CHEP 2019, 4-8 November, Adelaide Australia, AU
Accounting portal shows summary views of resource consumption (aggregated monthly)

Detailed accounting figures also available: DODAS pilot has consumed 760K CPU hours over 5 sites during October 2019

VOs active during October 2019
DODAS deployer has been used to exploit EGI compute resources; a successful test was done implementing a virtual site over 5 distinct providers:

- CMS Analysis Workflow used as demonstrator
  - Already supporting other communities

In general any DODAS supported use case can be deployed over EGI Federated Cloud:

- (Federated) HTCondor batch on demand, Spark, HDFS, Inference engine (TFaaS), etc.

See also: Dynamic integration of distributed, Cloud-based HPC and HTC resources using JSON Web Tokens and the INDIGO IAM Service

DODAS experts are available for consultancy and training interested communities.

- Contact: spiga@infn.it ; Additional information: Here
Thank you for your attention!

Questions?

eosc-hub.eu  @EOSC_eu