



Grid y Computación
de Altas Prestaciones

GRyCAP

Instituto de Instrumentación para
Imagen Molecular
Universitat Politècnica de València
Spain

ON-PREMISES SERVERLESS CONTAINER-AWARE ARCHITECTURE (OSCAR)

Alfonso Pérez, Sebas Risco, Germán Moltó,
Miguel Caballer, Amanda Calatrava

IBERGRID 2018

October 11-12, Lisbon, Portugal



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



MOTIVATION



- Scientific user with loads of data to process
 - Easily parallelizable
 - No access to specific cluster resources
 - No worries about configuration
- Easily configurable elastic cluster
 - Highly scalable
 - Easy to use
 - Upload a file as input
 - Download a file as output

GOALS OF THIS TALK



- What are Functions as a Service (FaaS) ?
- From Public to On-Premises FaaS
 - AWS Lambda
 - SCAR
 - OSCAR
- OSCAR use case:
 - Serverless plant recognition application

FROM IAAS TO FAAS



Private Cloud



Infrastructure
(as a service)



Platform
(as a service)





Function
(as a service)
(serverless arch)



Software
(as a service)

Private Cloud	Infrastructure (as a service)	Platform (as a service)	Function (as a service) (serverless arch)	Software (as a service)
Functions	Functions	Function	Functions	Functions
Data	Data	Data	Data	Data
Application	Data	Application	Application	Application
Runtime	Runtime	Runtime	Runtime	Runtime
Backend Code	Backend Code	Backend Code	Backend Code	Backend Code
OS	OS	OS	OS	OS
Virtualization	Virtualization	Virtualization	Virtualization	Virtualization
Server Machines	Server Machines	Server Machines	Server Machines	Server Machines
Storage	Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking	Networking

 Public Cloud Provider - responsibility

 Application Writer - responsibility

Awesome Vizualisation picked from : Ref : http://www.slideshare.net/manuel_silveyra/austin-cf-meetup-20150224/3

PUBLIC FAAS: AWS LAMBDA



- FaaS - Functions as a Service (Serverless Computing)
 - Executes functions in response to events with automated elasticity
- Stateless functions executed in closed environments
- Allows different programming languages:
 - Node.JS, Python, Java, C# and Go
- Pros
 - No infrastructure provision or configuration management
 - Large elasticity (thousands of parallel executions)
- Cons
 - Requires re-architecting the application
 - Difficult to customize execution environment



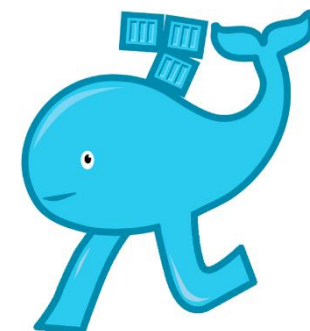
EXTENDING AWS LAMBDA: SCAR



GryCAP
Grid y Computación de Altas Prestaciones

www.grycap.upv.es

- Serverless Container-aware Architectures
 - Runs on AWS Lambda
 - Executes customized runtime environments provided by Docker containers (using uDocker)
 - Highly-parallel event-driven file-processing serverless applications
 - Bring your own application and runtime



grycap / scar

Unwatch ▾ 17

Unstar 316

Fork 23

Code

Issues 7

Pull requests 1

Projects 0

Wiki

Insights

Settings

Serverless Container-aware ARchitectures (e.g. Docker in AWS Lambda)

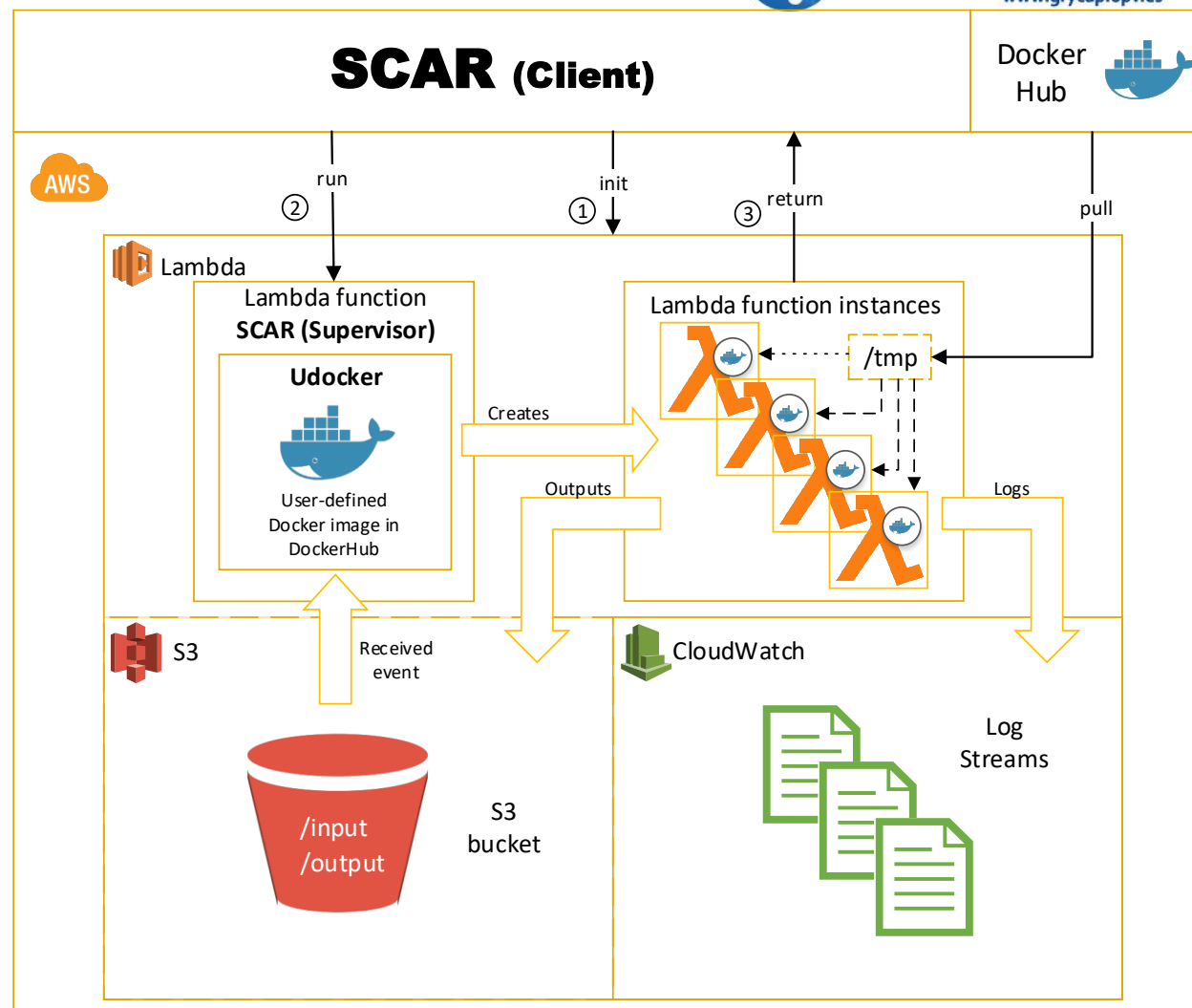
Edit

[docker](#) [aws-lambda](#) [serverless](#) [Manage topics](#)

SCAR: ARCHITECTURE

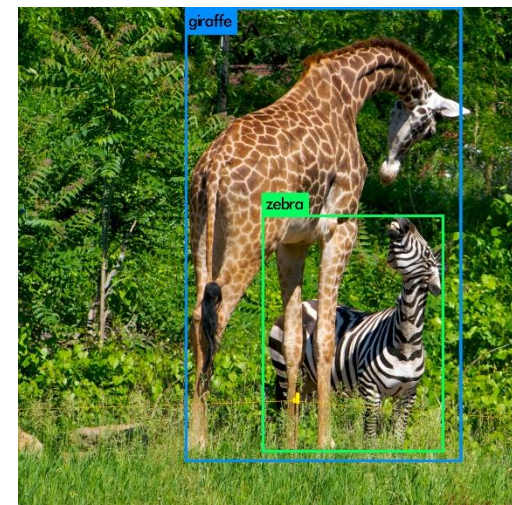


- SCAR Supervisor:
 - Retrieve the Docker image
 - Retrieve input file
 - Launch container
 - Store output file
- Useful for bursts of short stateless computational jobs





- Massive image recognition system using Neural Networks
 - Service triggered on-demand upon uploading a file into a bucket
- Supporting languages / applications not being easily supported by AWS Lambda
 - R, Erlang, Elixir
- Image / Video transformation



PUBLIC FAAS: LIMITATIONS



- Imposed by the public Cloud provider
- Size of the containers (< 230 MB)
- Execution time of the Jobs (~~5~~ **15** min)
- Solution
 - Create FaaS infrastructure in our own Cloud

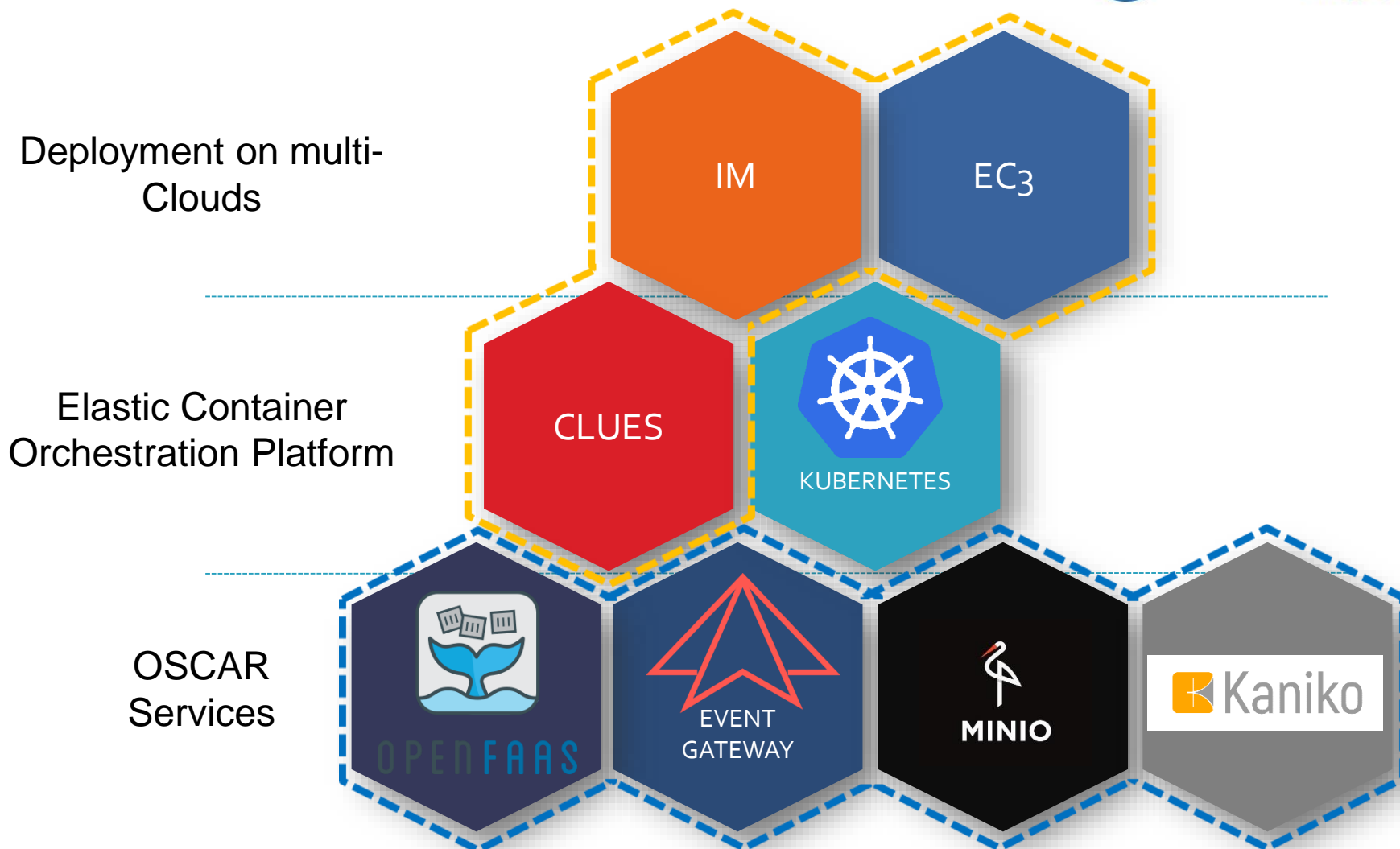
ON-PREMISES SCAR (OSCAR): COMPONENTS



GRyCAP

Grid y Computación de Altas Prestaciones

www.grycap.upv.es

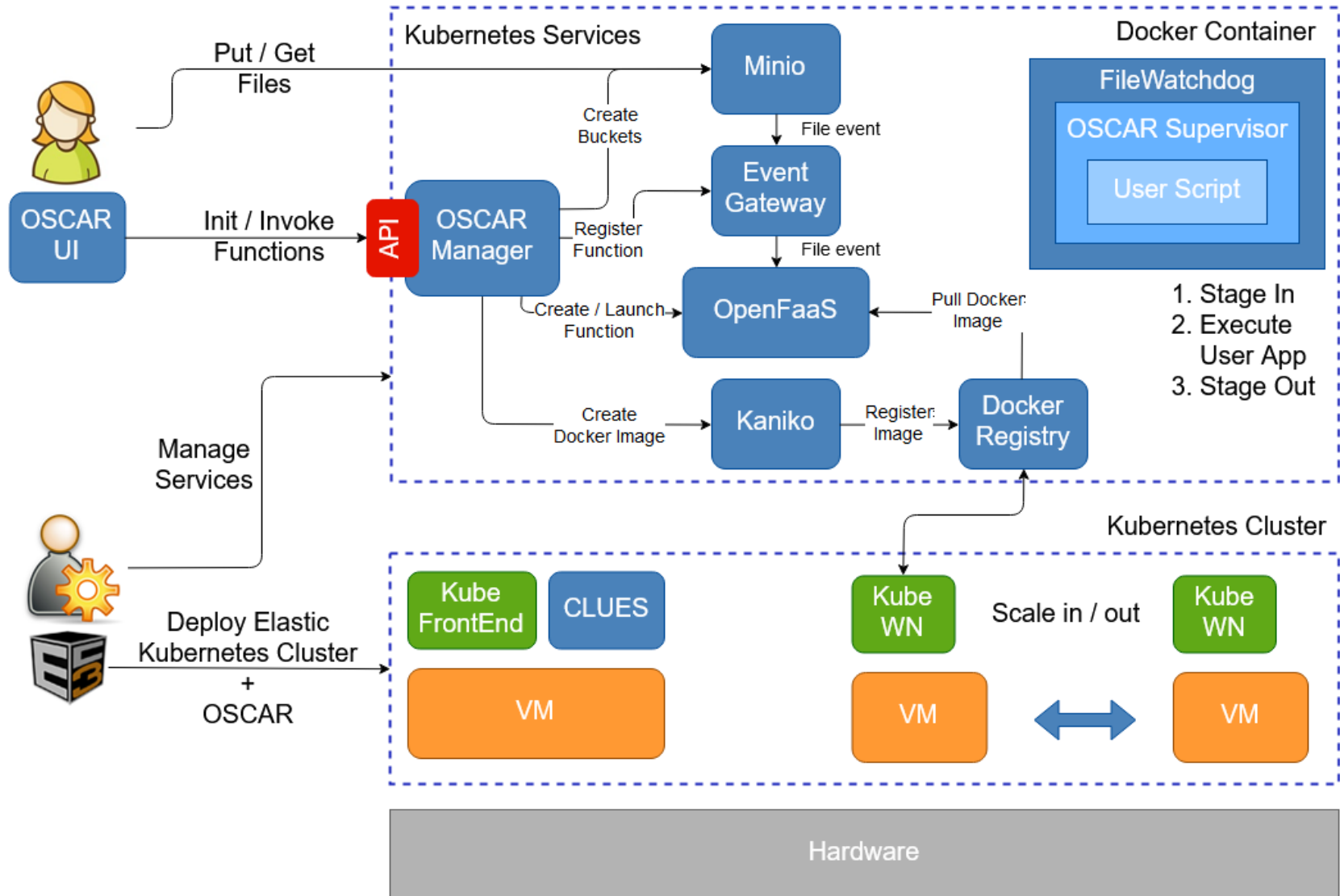


<http://github.com/grycap>

OSCAR ARCHITECTURE



On-Premises SCAR (OSCAR)





- Plant Classification with Lasagne/Theano
 - Example extracted from:
 - <https://github.com/indigo-dc/plant-classification-theano>
- Pretrained convolutional network
- Classifies a plant image among 6K plant species





swagger

http://158.42.105.19:32112/swagger.json

Authorize

Explore

On-premises Serverless Container-aware ARchitectures API Gateway

OSCAR API documentation

More documentation available on Github

<https://github.com/grycap/oscar>
[Apache2](#)

default

Show/Hide | List Operations | Expand Operations

POST	/function/async/{functionName}	Invoke a function asynchronously
GET	/function/{functionName}	Get a summary of an OpenFaaS function
POST	/function/{functionName}	Invoke a defined function
DELETE	/functions	Remove a deployed function.
GET	/functions	Get a list of deployed functions with: stats and image digest
POST	/functions	Deploy a new function.
PUT	/functions	Update a function.

DEMO TIME



POST /functions

Deploy a new function.

Parameters

Parameter	Value	Description	Parameter Type	Data Type
body	<pre>{ "image": "grycap/oscar-theano:latest", "name": "plant_test", "script": "IyEvYmluL2Jhc2gKCmVjaG8gIlNDUklQVDogSW52b2t1ZCBjbGFzc2lmeV9pbWFnZS5weS4gRmlsZSBhdmFpbGFibGUgaW4gJFNDQVJfSU5QVVRfRklMRSlKRklMRV90QU1FPWBiYXNlbnMtZSAkU0NB19JTBVVVF9GSUxhYApPVVRQVVRfRklMRTR0kU0NB19PVVRQVVRfRk9MREVSLyRGSUxhY05BTUUKCnB5dGhvb3NpZnlfZW50aW1hZ2UucHkgJFNDQVJfSU5QVVRfRklMRSA0byAkdT1VUUFVUX0ZJTEU=" }</pre>	Function to deploy	body	Model
				Example Value
				<pre>{ "image": "functions/nodeinfo:latest", "name": "nodeinfo", "script": "ZWNobyAiaXQgd29ya3MhIg==" }</pre>

Parameter content type:

Response Messages

HTTP Status Code	Reason	Response Model	Headers
202	Accepted		
400	Bad Request		
500	Internal Server Error		

Try it out!



Try it out!

[Hide Response](#)

Curl

```
curl -X POST --header 'Content-Type: application/json' --header 'Accept: text/plain' -d '{ \
  "image": "grycap/oscar-theano:latest", \
  "name": "plant-finder", \
  "script": "IyEvYmluL2Jhc2gKcmVjaG8gIlNDUk1QVDogSW52b2t1ZCBjbGFzc2lmeV9pbWFnZS5weS4gRmlsZSBhdmFpbGFibGUgaW4gJFNDQVJfSU5QVVRfRk1M
}' 'http://158.42.105.19:32112/functions'
```

Request URL

http://158.42.105.19:32112/functions

Response Body

no content

Response Code

202

Response Headers

```
{
  "content-length": "0",
  "content-type": "text/plain; charset=utf-8",
  "date": "Sun, 07 Oct 2018 20:41:13 GMT",
  "server": "Werkzeug/0.14.1 Python/3.7.0"
}
```

DEMO TIME



Minio Browser

🔍 Search Buckets...

📁 plant-finder-in

📁 plant-finder-out

🌐 158.42.105.19:31852

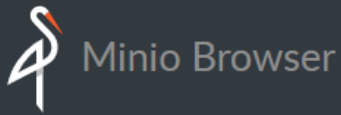
plant-finder-in /



Used: 74.50 MB



DEMO TIME



Search Buckets...

plant-finder-

plant-finder-

158.42.105.19

plant-finder-out /

Used: 74.50 MB



Predicted labels:

1. *Abies alba* | 68 %
2. *Picea abies* | 20 %
3. *Abies pinsapo* | 3 %
4. *Abies grandis* | 1 %
5. *Woodwardia radicans* | 1 %

CONCLUSIONS



- Functions as a Service frameworks provide a layer of abstraction on top of Container Orchestration Platforms
- In OSCAR
 - Elasticity of function invocations handled by the FaaS system
 - Elasticity of the Kubernetes cluster managed by CLUES and the IM
- High Throughput Computing Programming Model
- Future Work
 - Support for multiple data storage back-ends (EGI DataHub, dCache, etc.)
 - Web-based GUI
 - Hybrid (on-premises / public) workload distribution.

ACKNOWLEDGEMENTS



- The authors would like to thank the current projects BigCLOE^[1], DEEP Hybrid-DataCloud^[2] and the EGI Strategic and Innovation Fund for the financial support to continue the aforementioned developments.



[1] "Ministerio de Economía, Industria y Competitividad" for the project "*BigCLOE - Computación Big Data y de Altas Prestaciones sobre Multi- Clouds Elásticos*" with reference number TIN2016-79951-R.

[2] European Union's Horizon 2020 research and innovation programme under grant agreement No 777435.

CONTACT



GRyCAP
Grid y Computación de Altas Prestaciones

www.grycap.upv.es

Alfonso Pérez / Germán Moltó

Instituto de Instrumentación para Imagen Molecular

Universitat Politècnica de València

Camino de Vera s/n

46022, Valencia

SPAIN



alpegon3@upv.es

gmolto@dsic.upv.es

<http://www.grycap.upv.es/gmolto>

 @gmolto