



Review SME Continuous Call (F4Fp – SME) FRIDA – Stage 1

Celestino Alvarez
Teiichi Capital Tech.

Virtual FEC 8
The World, November 13th 2020



Outline

FINANCIAL RESEARCH INVOLVING DATA ANALYSIS – FRIDA – STAGE 1

- **Experiment description (max. 4 slides)**
 - Concept and objectives
 - Background and motivation
 - Experiment set-up
- **Project results (max. 3 slides)**
 - Measurements
 - Lessons learned
- **Business impact (min. 4 slides)**
 - Impact on your business, .. how did Fed4FIRE helped you ?
 - Value perceived, .. why did you come to Fed4FIRE ?
- **Feedback (min. 4 slides)**
 - Used resources and tools
 - Added value of Fed4FIRE



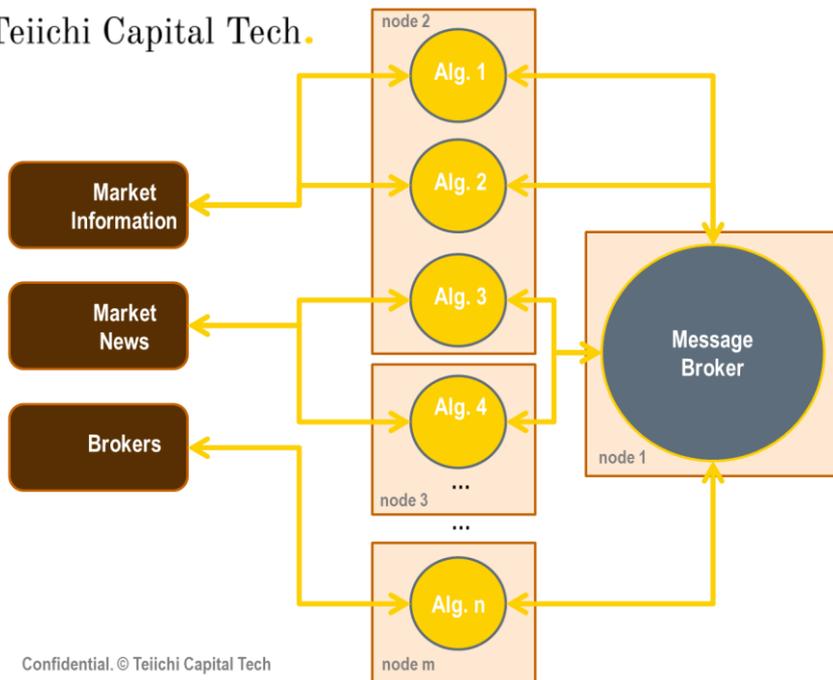
Experiment Description

FINANCIAL RESEARCH INVOLVING DATA ANALYSIS – FRIDA – STAGE 1



Concept and Objectives

Teiichi Capital Tech.



Confidential. © Teiichi Capital Tech

Objectives

Stage 1

O1. Measure and compare the performance of the message broker and network while scaling up the volume of messages processed

Stage 2

O2. Improve and automate the machine learning workflow for training, evaluating and deploying new algorithms

O3. Dimension the infrastructure required for scaling up

Background and Motivation

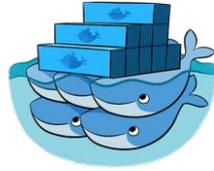
Teiichi Capital Tech is a fintech that has developed a distributed cloud platform for processing financial data, taking decisions and executing operations in real time, fully automated, using cutting edge AI algorithms. The platform is currently oriented to ingest data from stock markets, seeking deals to obtain +20% monthly ROI.

The company intends to **scale the platform** to process up to +16.000 financial instruments, so it is needed to **optimize and validate the evolution of the platform**, which has a critical impact in our business.

Experiment Setup



AKIRA



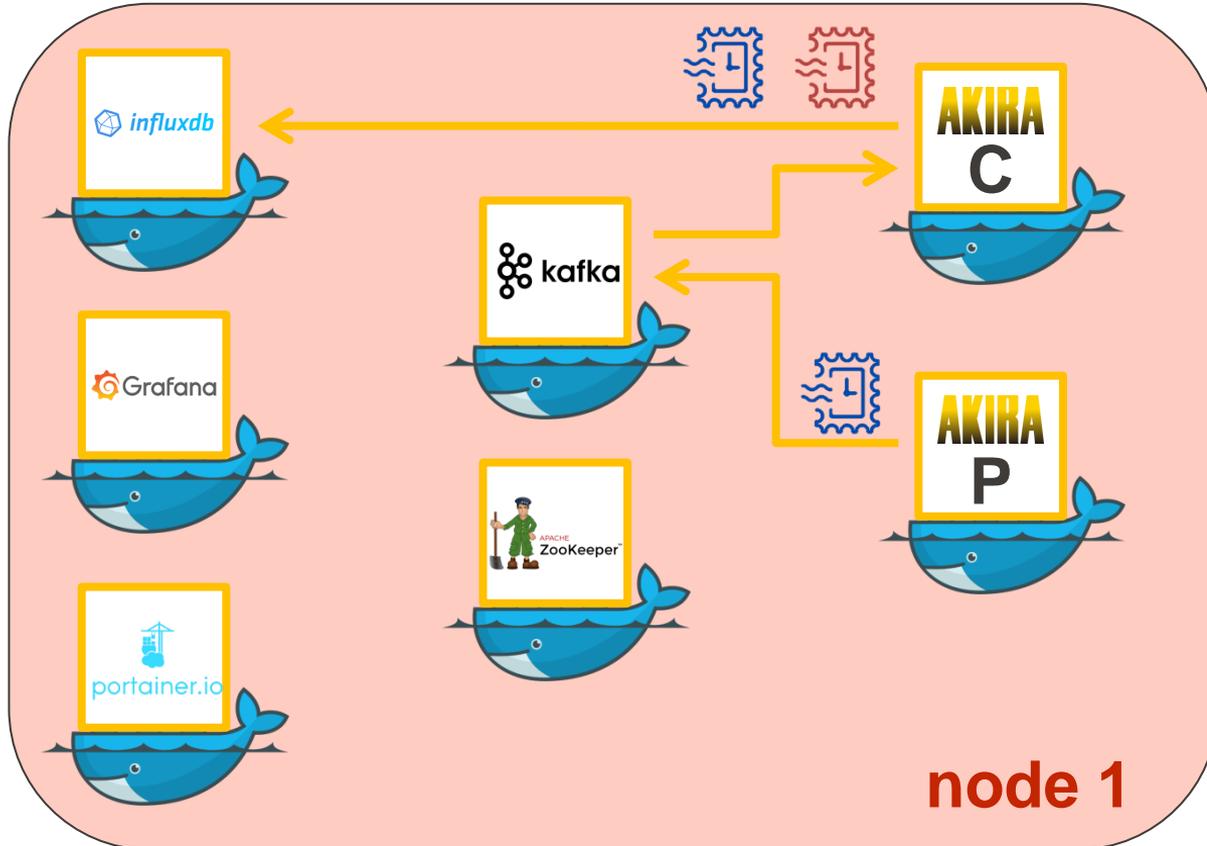
Name	Stack	Image	Scheduling Mode	Port
portainer_agent	portainer	portainer/agent.latest	global 2 / 2	-
portainer_cadvisor	portainer	gcr.io/google-containers/cadvisor.latest	replicated 3 / 3 1 Scale	877
portainer_collectd	portainer	puskel/docker-collectd.latest	global 2 / 2	-
portainer_consumer	portainer	teichu/consumer.latest	replicated 3 / 300 100	-
portainer_grafana	portainer	grafana/grafana.latest	replicated 3 / 1 1 Scale	871
portainer_influxdb	portainer	influxdb/latest	replicated 3 / 1 1 Scale	874
portainer_kafka	portainer	kafka-docker_kafka.latest	replicated 3 / 1 1 Scale	879
portainer_kafka_manager	portainer	niebaub/kafka-manager.stable	replicated 3 / 3 1 Scale	877
portainer_portainer	portainer	portainer/portainer.latest	replicated 3 / 3 1 Scale	874
portainer_producer	portainer	teichu/producer.latest	replicated 336 / 336 1 Scale	-



```
malvarez@fsophia:~$ ./register_and_get_ip.sh ". /worker_node.sh
docker_token_1598537715209158862 172.16.132.9
Reserving node..
Reserved with IP: 172.16.132.26
```

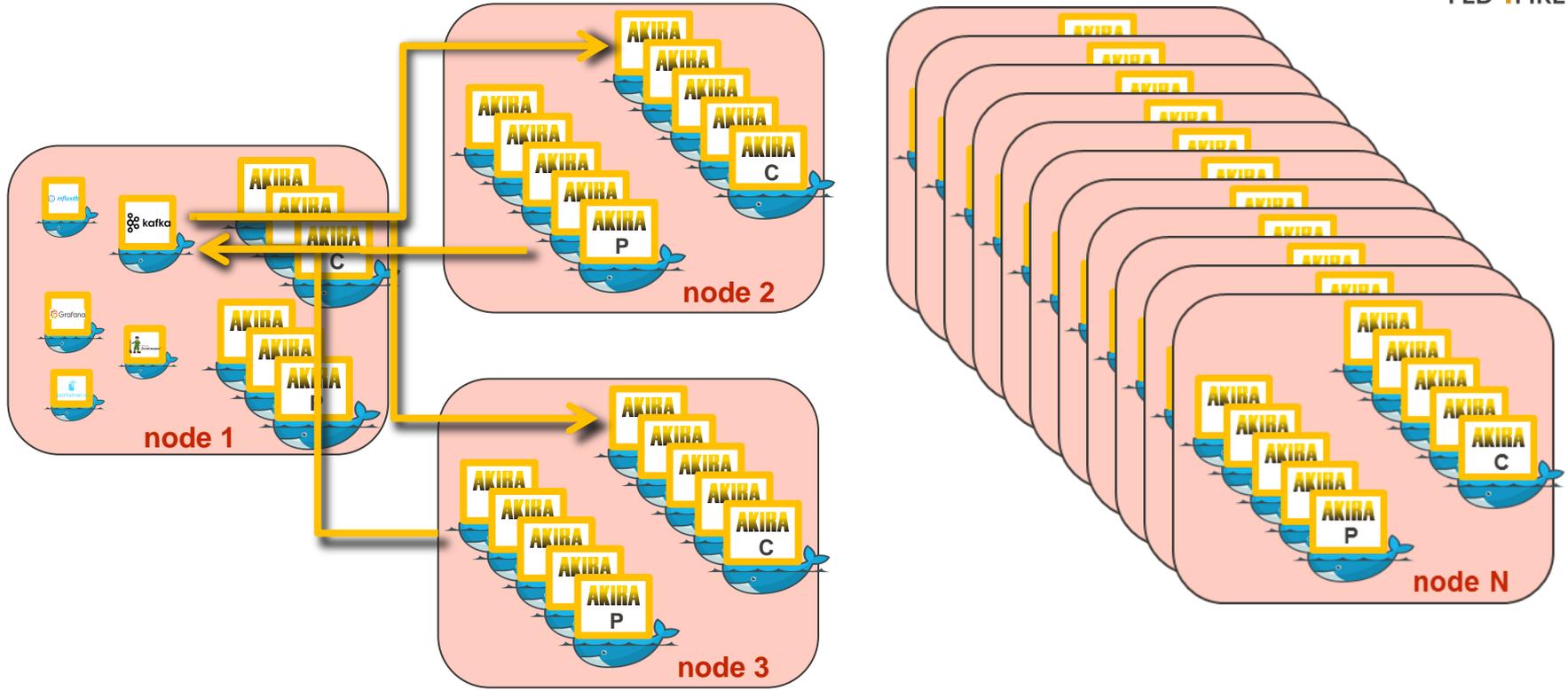


Experiment Setup @sophia ubv @grid5000



1. Deploy the stack in the swarm
2. Add more nodes
3. Scale containers / services

Experiment Setup @sophia ubv @grid5000



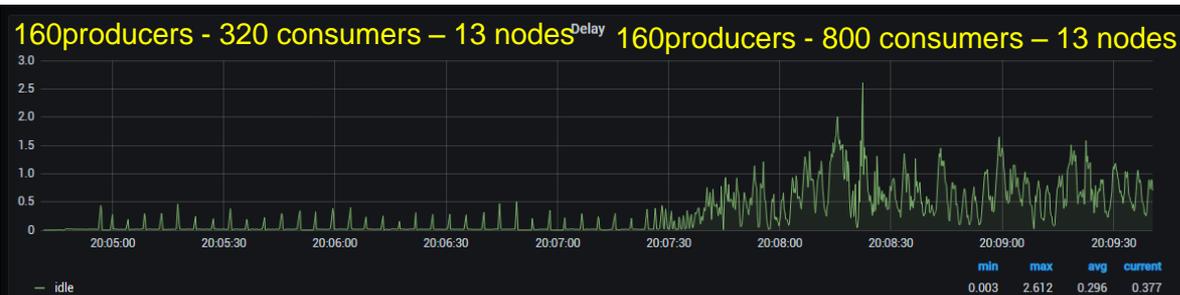
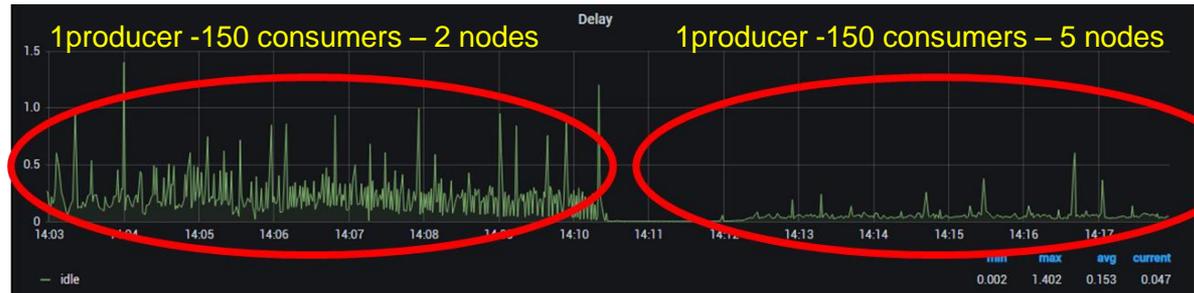


Project Results

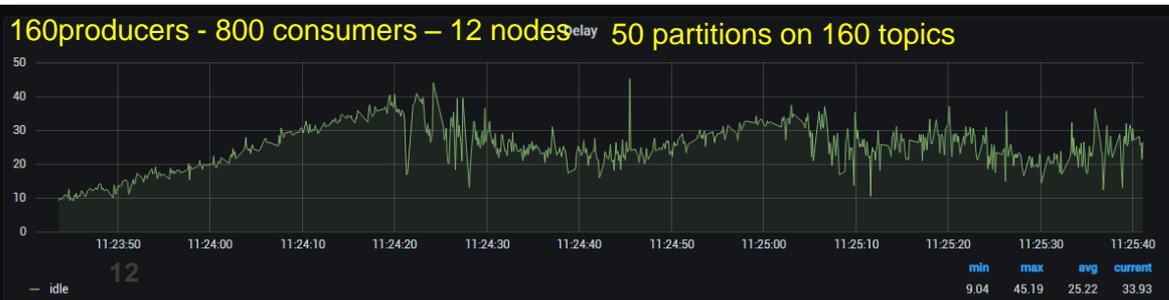
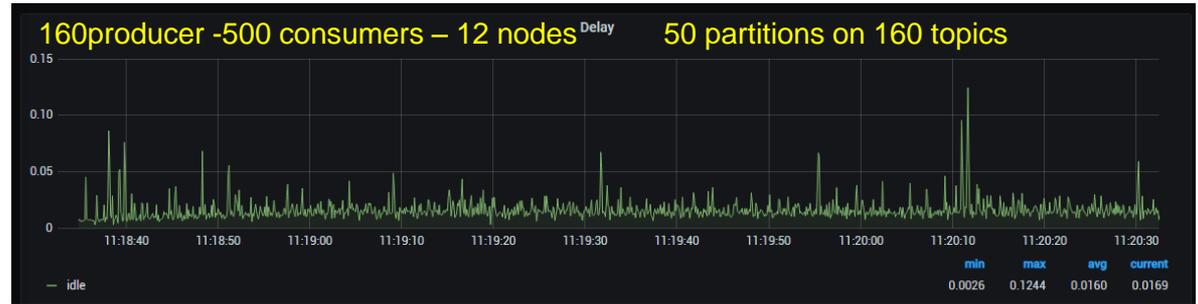
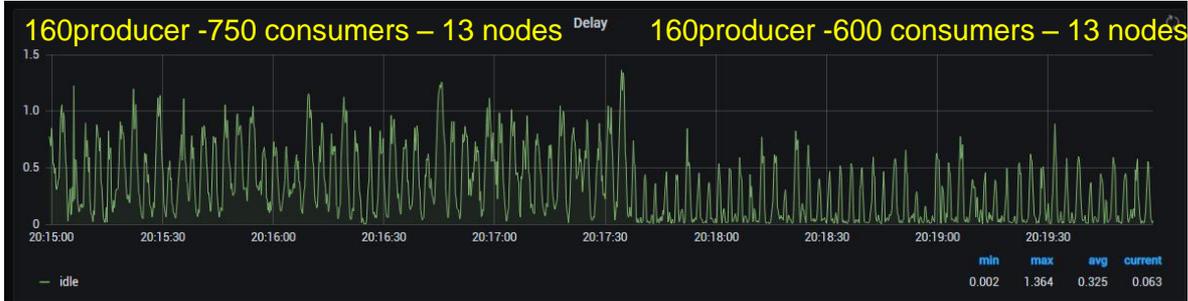
FINANCIAL RESEARCH INVOLVING DATA ANALYSIS – FRIDA – STAGE 1



Measurements



Measurements



Lessons Learned

- Learn about G5K and tools (OAR specifically)
- Hands on experience with Docker swarm
- NTP do not provide enough accuracy for this kind of experiment
- Docker swarm fits perfectly for managing the scalability of our platform
- Kafka is a very fast and reliable server for our purpose
- Increasing the number of producers does not affect the performance of a server.
- For a single Kafka server, 500-600 producers are the limit for a proper delivery time.
- Fine tuning of the server is required to increase performance and a Kafka servers cluster maybe required.



Business Impact

FINANCIAL RESEARCH INVOLVING DATA ANALYSIS – FRIDA – STAGE 1



Impact on your business

- Scaling the business depends on scaling the technology
 - Need to go from 4-5 assets to +16.000 assets
- Early technology decision have a strong impact on the future business, but **how to take this decisions being an SME with limited resources?**
 - Take for granted the product documentation and cross your fingers
 - Use Fed4FIRE infrastructure to validate your use case

Impact on your business

- After the experiment we have obtained hands-on experience on how to scale a core component of our platform
- We are confident on technical decisions we have made and we expect the platform will be able to handle up to our targeted of 160M Eur of assets reliably.
- Now we have clear metrics to dimension part of the platform to scale on demand.

Value Perceived

- Availability of a large amount of resources
 - G5000 is a great infrastructure that can be easily used to test different software possibilities
- Funding helps to lower the cost of learning curve of the platform
 - But furthermore, it allows to dedicate time to experiment, which is important but not urgent in an SME, where costs require to be assigned to billable projects

Value Perceived

- Being backed by Fed4FIRE it is possible to target more ambitious projects
 - Being able to use those resources and the experience of the patrons will enable us to test i.e. a lambda architecture for a continuous ML workflow, including automatic deployment to production environment. This is the objective of the Stage 2 project.



Feedback

FINANCIAL RESEARCH INVOLVING DATA ANALYSIS – FRIDA – STAGE 1



Used resources and tools

- Testbed: Grid5000
 - At the time we started the experiment there were no integration with Jfed, so we used G5K credentials. For Stage2 we will be using it for connecting Tengu/GPULab with lyon.g5k cluster.
- From the different sites available, we choose Sophia, as we feel it was not very used (probably because older machines were there) but were enough for our experiment.

Used resources and tools

- We found issues with NTP

```
malvarez@uvb-43:~$ ntpq -c 'rv 0 offset'  
offset=-0.676412
```

```
malvarez@uvb-9:~$ ntpq -c 'rv 0 offset'  
offset=-0.898149
```

- Solution: use ptpd

```
sudo-g5k apt-get install -y -qq --no-install-recommends ptpd >/dev/null  
sudo-g5k service ntp stop  
sudo-g5k ptpd -m -i br0 &
```

```
# on worker nodes  
sudo-g5k ptpd -i br0 &
```

Used resources and tools

- Quickly learn on how to use OAR for reserving nodes. Good tutorials and reference sheets.
- Tried to use ENOS but following the tutorial produced some errors:

```
TASK [Gathering Facts] *****
ok: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : include] *****
included: /home/malvarez/enos/venv/lib/python3.7/site-packages/enos/ansible/roles/influx/tasks/deploy.yml for uvb-38-kavlan-4.sophia.grid5000.fr

TASK [influx : Checking that the vip is free on this host] *****
changed: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : Add a vip address for influx db] *****
skipping: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : Copying over the influx conf] *****
ok: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : Copy types.db] *****
ok: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : Start the influx container] *****
changed: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : Waiting for the influx service to become available] *****
ok: [uvb-38-kavlan-4.sophia.grid5000.fr]

TASK [influx : Create the cadvisor database] *****
fatal: [uvb-38-kavlan-4.sophia.grid5000.fr]: FAILED! => ("changed": false, "module_stderr": "Shared connection to uvb-38-kavlan-4.sophia.grid5000.fr closed.\n", "module_stdout": "Traceback (most recent call last):\n  File \"/root/.ansible/tmp/ansible-tmp-1589984142-9992378-247239009468819/ansible_influxdb_database.py", line 18, in <module>\n    ansible_main()\n  File \"/root/.ansible/tmp/ansible-tmp-1589984142-9992378-247239009468819/ansible_influxdb_database.py", line 105, in ansible_main\n    invoke_module(zipped_mod, temp_path, ANSIBALLZ_PARAMS)\n  File \"/root/.ansible/tmp/ansible-tmp-1589984142-9992378-247239009468819/ansible_influxdb_database.py", line 48, in invoke_module\n    imp.load_module('__main__', mod, module, MOD_DESC)\n  File \"/tmp/ansible_influxdb_database_payload_f03m5/_main.py", line 143, in <module>\n    File \"/tmp/ansible_influxdb_database_payload_f03m5/ansible_influxdb_database_payload_f03m5/module_utils/influxdb.py", line 26, in connect_to_influxdb\n    File \"/usr/local/lib/python2.7/dist-packages/influxdb/client.py", line 118, in __init__\n    self._udp_port = int(udp_port)\nTypeError: int() argument must be a string or a number, not 'NoneType'\n", "msg": "MODULE FAILURE\nSee stdout/stderr for the exact error", "rc": 1}
to retry, use: --limit @/home/malvarez/enos/venv/lib/python3.7/site-packages/enos/ansible/enos.retry

PLAY RECAP *****
uvb-25-kavlan-4.sophia.grid5000.fr : ok=16  changed=3  unreachable=0  failed=0
uvb-33-kavlan-4.sophia.grid5000.fr : ok=16  changed=3  unreachable=0  failed=0
uvb-38-kavlan-4.sophia.grid5000.fr : ok=28  changed=6  unreachable=0  failed=1

('code': 2, 'result': [{'uvb-25-kavlan-4.sophia.grid5000.fr': {'ok': 16, 'failures': 0, 'unreachable': 0, 'changed': 3, 'skipped': 1}}, {'uvb-33-kavlan-4.sophia.grid5000.fr': {'ok': 16, 'failures': 0, 'unreachable': 0, 'changed': 3, 'skipped': 1}}, {'uvb-38-kavlan-4.sophia.grid5000.fr': {'ok': 28, 'failures': 1, 'unreachable': 0, 'changed': 6, 'skipped': 3}}], 'playbook': '/home/malvarez/enos/venv/lib/python3.7/site-packages/enos/ansible/enos.yml')
ERROR: ansible-playbook: Failed hosts: [uvb-38-kavlan-4.sophia.grid5000.fr]
```

Used resources and tools

- Alternatively, we tested Docker swarm and we feel it was the right way to go.
 - We create some bash scripts based on OAR to create the swarm and add more nodes.

```
malvarez@fsophia:~$ ./launch_test.sh
```

```
malvarez@fsophia:~$ more docker_token_1598537715209158862  
SWMTKN-1-1wwm3ag2nm6t051zps037n43an44fsu9lnmvbe040xbb7qyosp-7pnu3ieffymtyca18952394pn
```

```
malvarez@fsophia:~$ ./register_and_get_ip.sh "./worker_node.sh"  
docker_token_1598537715209158862 172.16.132.9  
Reserving node..  
Reserved with IP: 172.16.132.26
```

Used resources and tools

- G5000 usage policy is quite complex (quite a lot of rules)
 - When you are focused on your experiment is sometimes difficult to have it in mind and we broke some rules. (sorry for that!)
- Having ptpd, influxdb and grafana ready to be used, could be a good suggestion for similar measurements as ours.
 - We are happy to contribute with our scripts and Dockerfiles

Added value of Fed4FIRE

- Having such amount of resources ready to be used is the most appreciated value. Also the good documentation to get hands on quickly is a plus.
- In Stage 2 we will take benefit of different testbeds and will be able to compare different philosophies as in GPULab and Grid5000.





Co-funded by the
European Union



Co-funded by the
Swiss Confederation

This project has received funding from the European Union's Horizon 2020 research and innovation programme, which is co-funded by the European Commission and the Swiss State Secretariat for Education, Research and Innovation, under grant agreement No 732638.

WWW.FED4FIRE.EU