Inria
inventeurs du monde numérique

github.com/scampion/orqal/

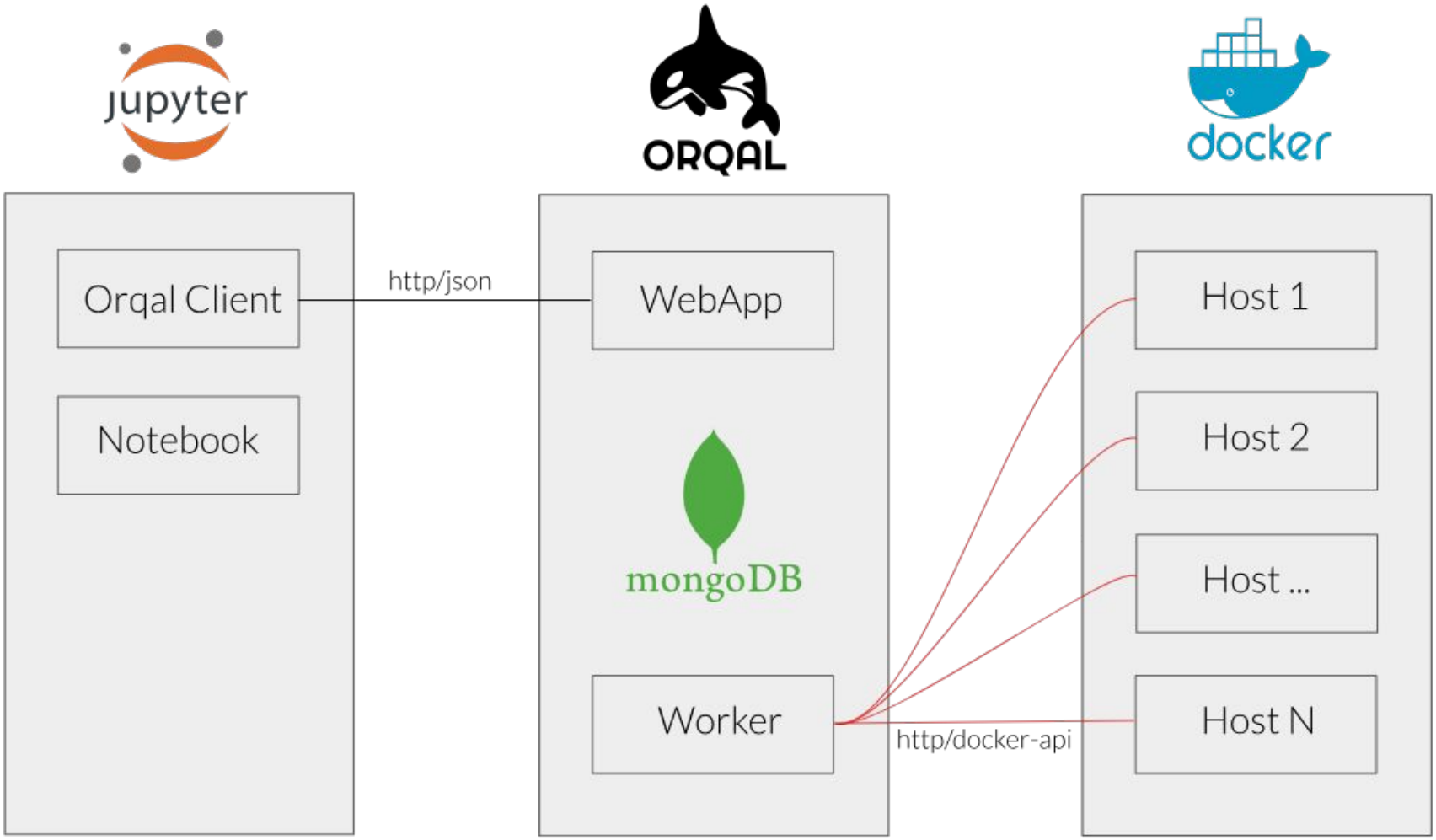
ORQAL Beta

Orqal for [ORQ]chestration of [AL]gorithms is a simple batch scheduler for docker cluster which can be used remotely and without overhead in scientific experiment.

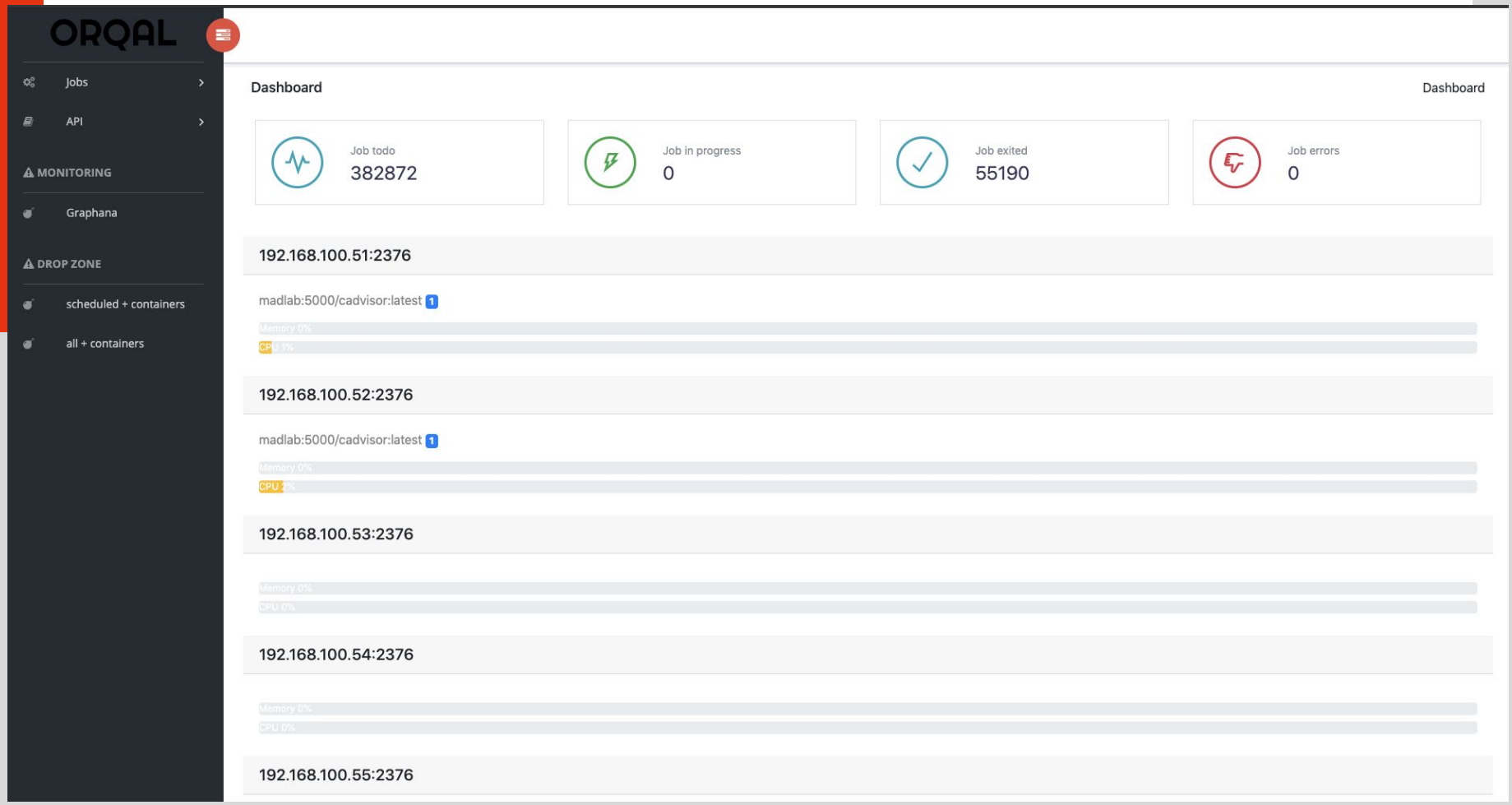
- Drive a cluster of docker nodes
- HTTP/API
- best fit scheduling
- Python client, Go shortly ;)

Goals

- Schedule jobs & share ressources
- Run large batch of jobs
- Fit my user experience, I'm working on my laptop in my notebook
- Reproduce
- Robust
- Experimentation recovery
- Switch easily from local job to distributed
- Easy to admin and maintain



Dashboard



How to deploy a new service

1. Push your container

```
docker push madlab:5000/myawesometool
```

2. Write an Orgal wrapper

```
class Rabin2(AbstractWrapper):
    docker_url = "madlab:5000/radare2"
    volumes = {'/database': {'bind': '/database', 'mode': 'ro'}}
    threads = 1
    memory_in_gb = 1

    def get_cmd(self, params):
        return "rabin2 -I %s" % self.job.input

    def set_result(self, job):
        r = {l.split()[0].replace('.', '_'): l.split()[1] for l in job.stdout if len(l.split()) == 2}
        job.set_result(r)
```

In practice @ INRIA ...

Hardware - UCS

204 Gflops
2.3 TB of RAM

Management Servers - 3 nodes :
- 8 cores
- 64GB of RAM

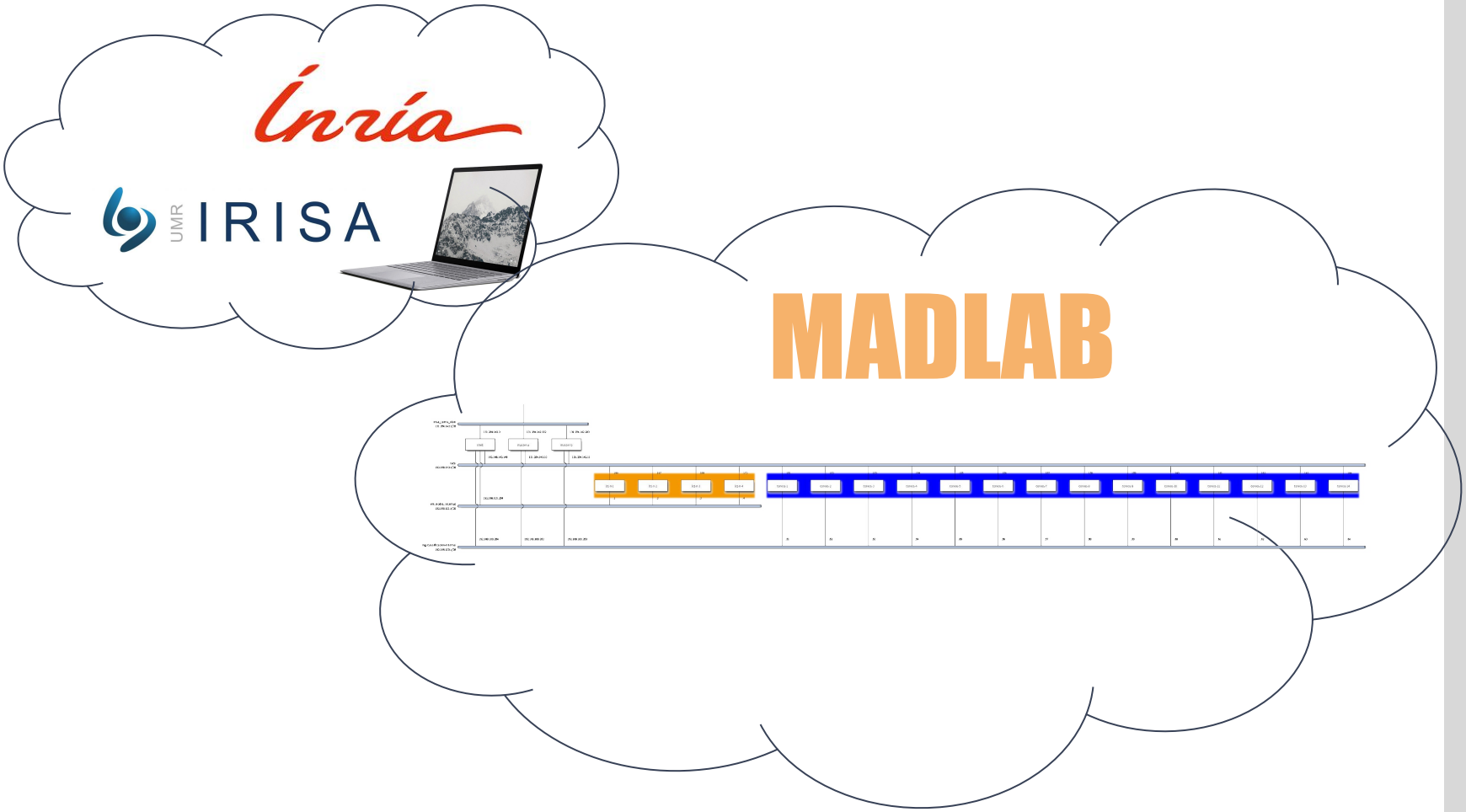
Computing Clusters - 18 nodes :
- 64 cores
- 128 GB of RAM

Software / Services

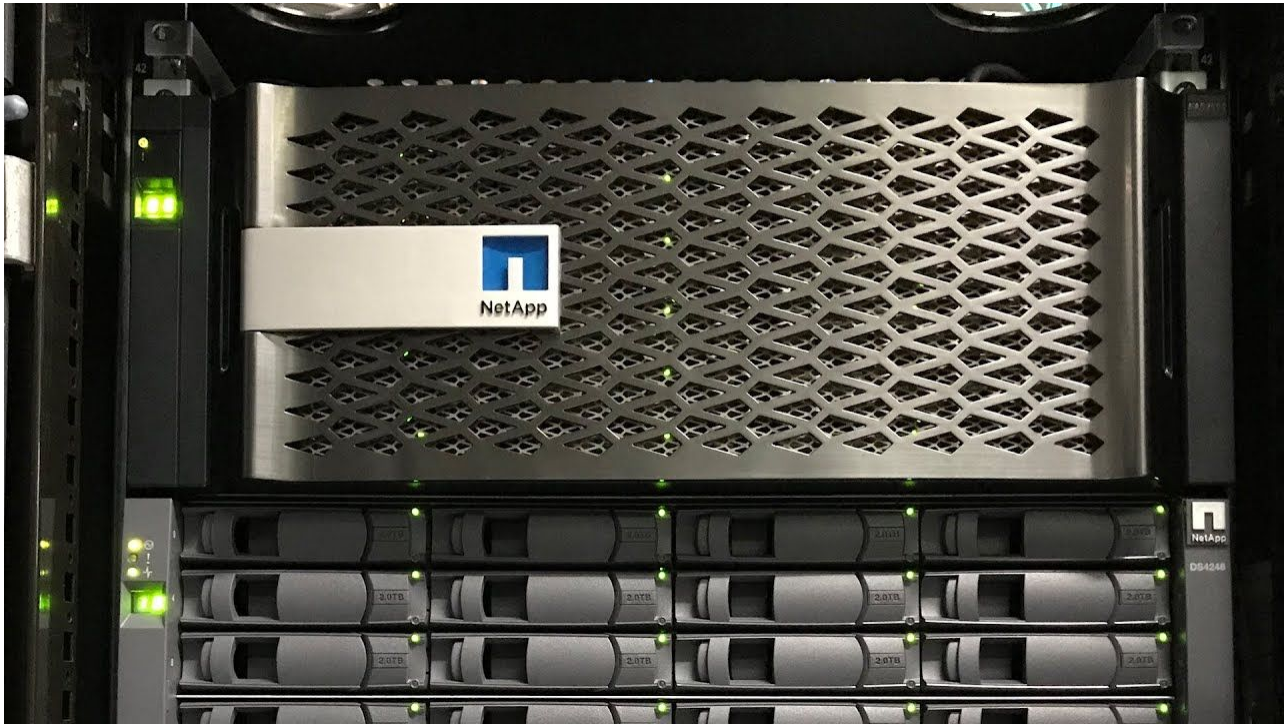
- The paradigm : 1 file + 1 algo = 1 result
- Stateless
- Services Availables :
  - **Rabin2**
  - **SCDG Extraction**
  - **...**

- Typical experiment use case:
  - 1. Build a dataset using a query
    - 'I need x86 and windows binary' > 120 000 files
  - 2. Run feature extraction on it :
    - for each b in x86_win_binaries:
      - job(b, service=A, params= { alpha = 0,3 ; beta = 0.6 ; ….})
  - 3. Wait jobs
  - 4. Explore results

Hardware - Private Network



Hardware - Storage



NetApp Storage
16 TB

/database
/scratch

/database - Read Only Datasets

- annotated_malwares
- babic_database_categorized
- babic_database_gs.json
- babic_database_scdg
- binaries.db
- cleanware.kch
- conference_malwares
- demo_database
- demo_database_gspan.json
- demo_database_gspan_oriented.json
- gspan_test_dataset
- linux_static_cleanwares
- malware_feed (438 000 binaries)
- microsoft_challenge
- mirai
- mirai_database_v1
- mirai_graphs
- mirai_graphs.zip
- mirai_samples.db
- nodes_status
- obfuscated_linux_static_cleanwares
- obfuscated_mirais
- packed
- unpacking_development_dataset
- virus_total_samples

/scratch - 2 Month Max

- Dedicated to experimental results
- mkdir /scratch/login
- Life cycle /!\ 2 Months max

```
find /scratch/ -type f -mtime +60 -exec rm {} \;
```