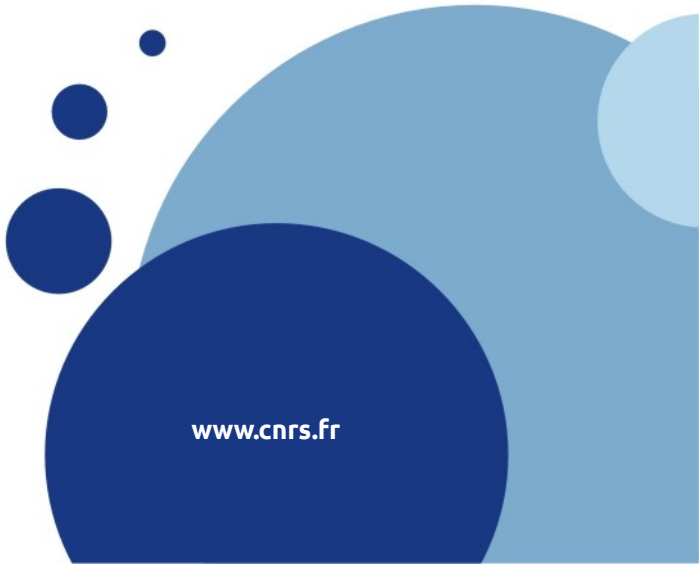
How to set up a Galaxy service you can count upon

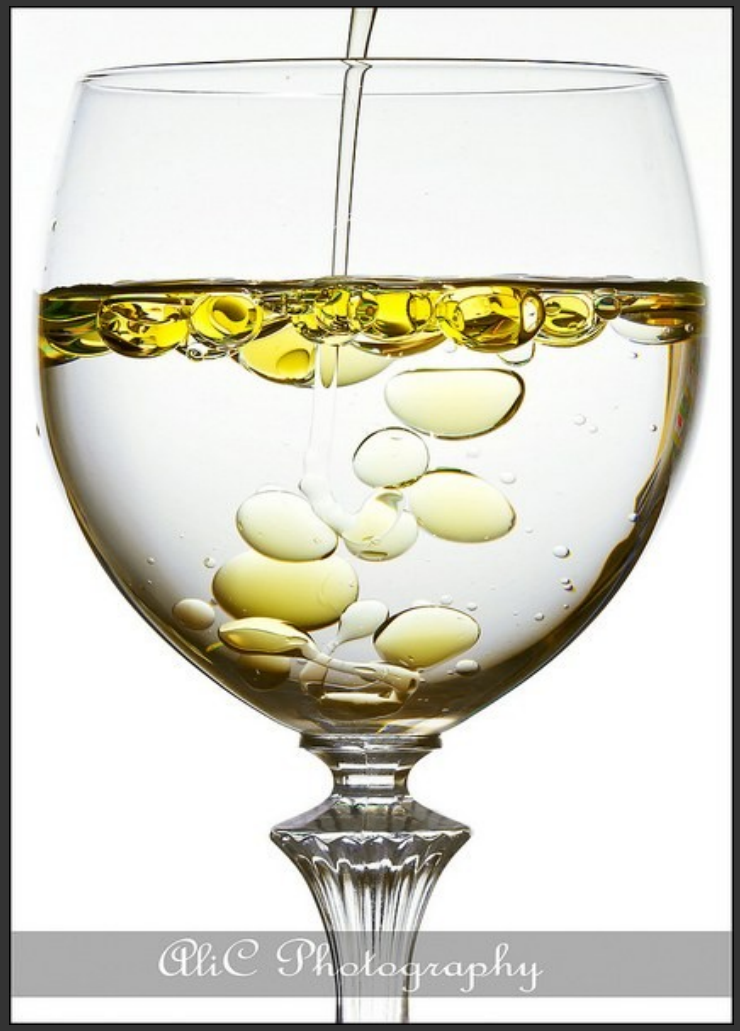
by Vladimir **Darić**

Institut de Génétique et Microbiologie, UMR 8621



The Zen of Galaxy

- Keep your modifications isolated
- Use virtualenv
- Test
- Test more !
- Use scalable storage/cluster
- Know what Galaxy is doing



Don't mix your code & configuration with Galaxy

HOW TO :

- Your xml configuration goes here :

  universe_wsgi.ini :

  tool_config_file=tool_conf.xml,shed_tool_conf.xml,my_own_confs.xml

- Your app's folder structure :

  tools/

  myApp/

  myapp.py

  myapp.xml

- MyApp → my_own_conf.xml

Ref: https://wiki.galaxyproject.org/GalaxyToolPanel

Use *virtualenv*

- Virtualenv : for isolated *Python* environments

- Control your libraries
- Each virtualenv has its own pip/easy_install
- pip freeze > requirements.txt

Refs : https://wiki.galaxyproject.org/Admin/Config/Performance/ProductionServer
http://www.virtualenv.org/en/latest/

Test !

- run_unit_tests.sh
- run_functional_tests.sh
- run_tool_shed.sh
- run_tool_shed_functional_tests.sh

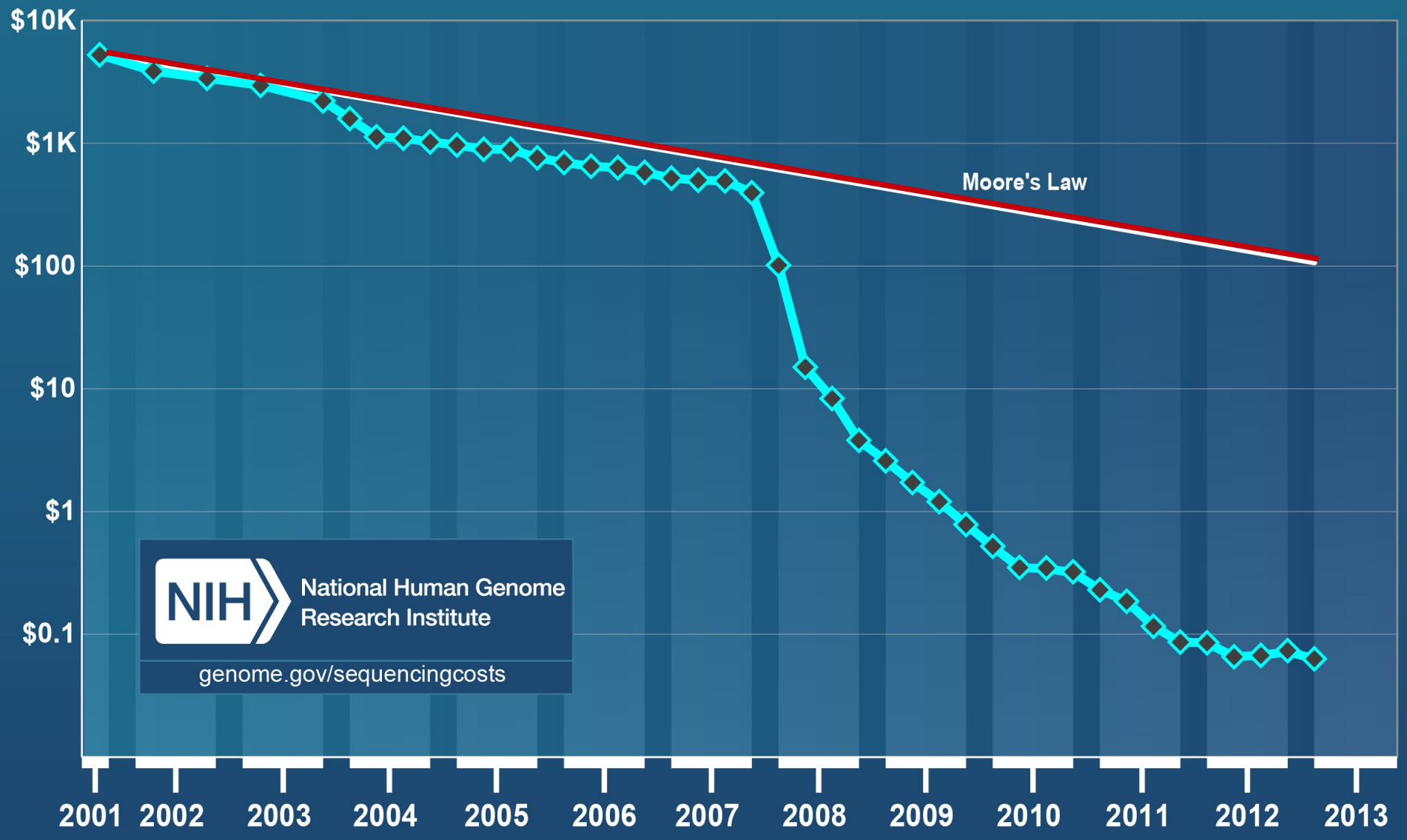
Ref : https://wiki.galaxyproject.org/Admin/Running%20Tests

• Use scalable storage/cluster

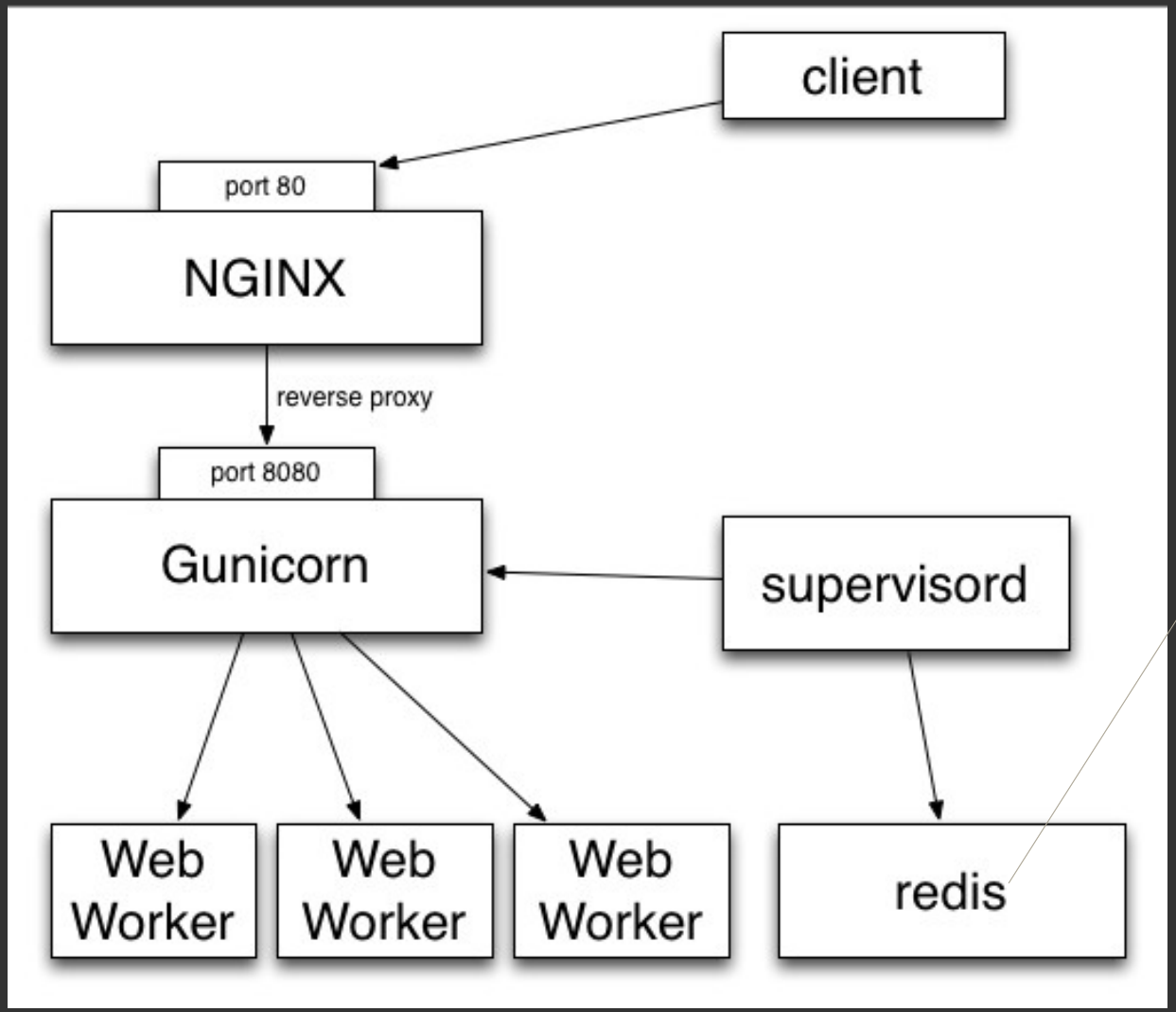
- ZFS rocks !
  - some theoretical limits in ZFS are :
    - 2^{48}: number of entries in any individual directory
    - 16 Exbibytes (2^{64} bytes): maximum size of a single file
    - 16 Exbibytes: maximum size of any attribute
    - 256 Zebibytes (2^{78} bytes): maximum size of any zpool
    - 2^{56}: number of attributes of a file (actually constrained to 2^{48} for the number of files in a ZFS file system)
    - 2^{64}: number of devices in any zpool
    - 2^{64}: number of zpools in a system
    - 2^{64}: number of file systems in a zpool
  - Snapshots, deduplication, compression & more

Ref : http://en.wikipedia.org/wiki/Zfs

Cost per Raw Megabase of DNA Sequence



Web stack the python way (The GIL issue)



REmote DIctionary Server

Dans Galaxy : universe_wsgi.ini

- Sections :
  - server (web)
  - ~~Manager~~
  - Handler

```
[server:web0]
use = egg:Paste#http
port = 8080
host = 127.0.0.1
use_threadpool = true
threadpool_workers = 7

[server:web1]
use = egg:Paste#http
port = 8081
host = 127.0.0.1
use_threadpool = true
threadpool_workers = 7
```

```
[server:handler0]
use = egg:Paste#http
port = 8090
host = 127.0.0.1
use_threadpool = true
threadpool_workers = 5

[server:handler1]
use = egg:Paste#http
port = 8091
host = 127.0.0.1
use_threadpool = true
threadpool_workers = 5
```

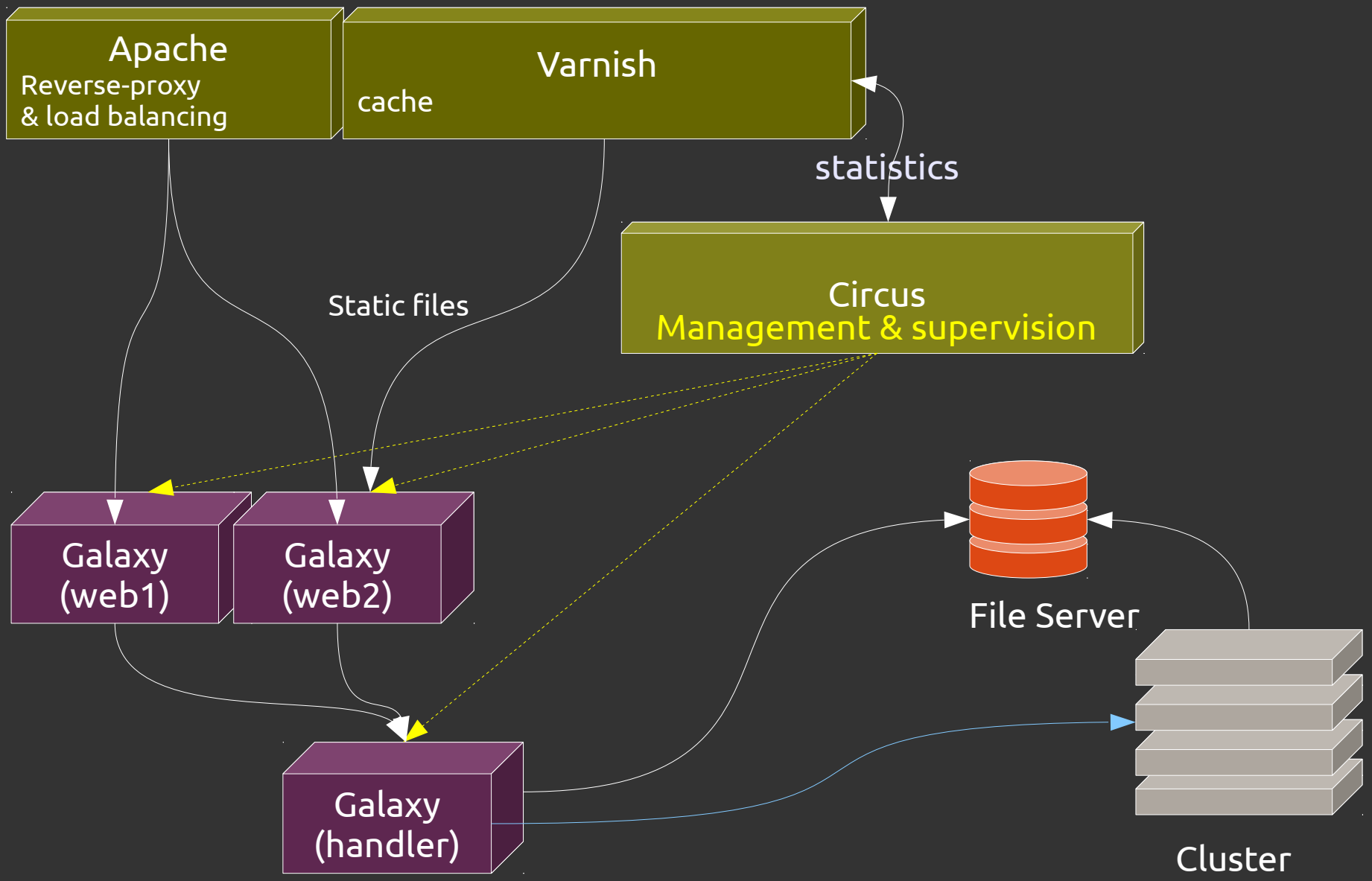
https://wiki.galaxyproject.org/Admin/Config/Performance/Scaling

Monitor your Galaxy

- What is Circus ?
  - process manager
    - Starts & stops processes
    - Automatic « respawn », contrôle du « flapping »
  - monitoring
  - Real time statistics
    - ZeroMQ → asynchronous communication

Ref : http://circus.readthedocs.org/en/latest/

Notre instance Galaxy



Logout

Add Watcher

- HOME
- OVERVIEW
- OPTIONS
- HELP

Watchers

Name Status	Processes	Command	Shell	uid	gid
handler0	1	python ./scripts/paster.py serve universe_wsgi.ini –server-name=handler0 –pid-file=/srv/galaxy/galaxy-dist/run/galaxy-handler0.pid –log-file=/srv/galaxy/galaxy-dist/log/galaxy-handler0.log	False	None	None
main	1	python ./scripts/paster.py serve universe_wsgi.ini –server-name=main –pid-file=/srv/galaxy/galaxy-dist/run/galaxy-main.pid –log-file=/srv/galaxy/galaxy-dist/log/galaxy-main.log	False	None	None
manager	1	python ./scripts/paster.py serve universe_wsgi.ini –server-name=manager –pid-file=/srv/galaxy/galaxy-dist/run/galaxy-manager.pid –log-file=/srv/galaxy/galaxy-dist/log/galaxy-manager.log	False	None	None
web1	1	python ./scripts/paster.py serve universe_wsgi.ini –server-name=web1 –pid-file=/srv/galaxy/galaxy-dist/run/galaxy-web1.pid –log-file=/srv/galaxy/galaxy-dist/log/galaxy-web1.log	False	None	None

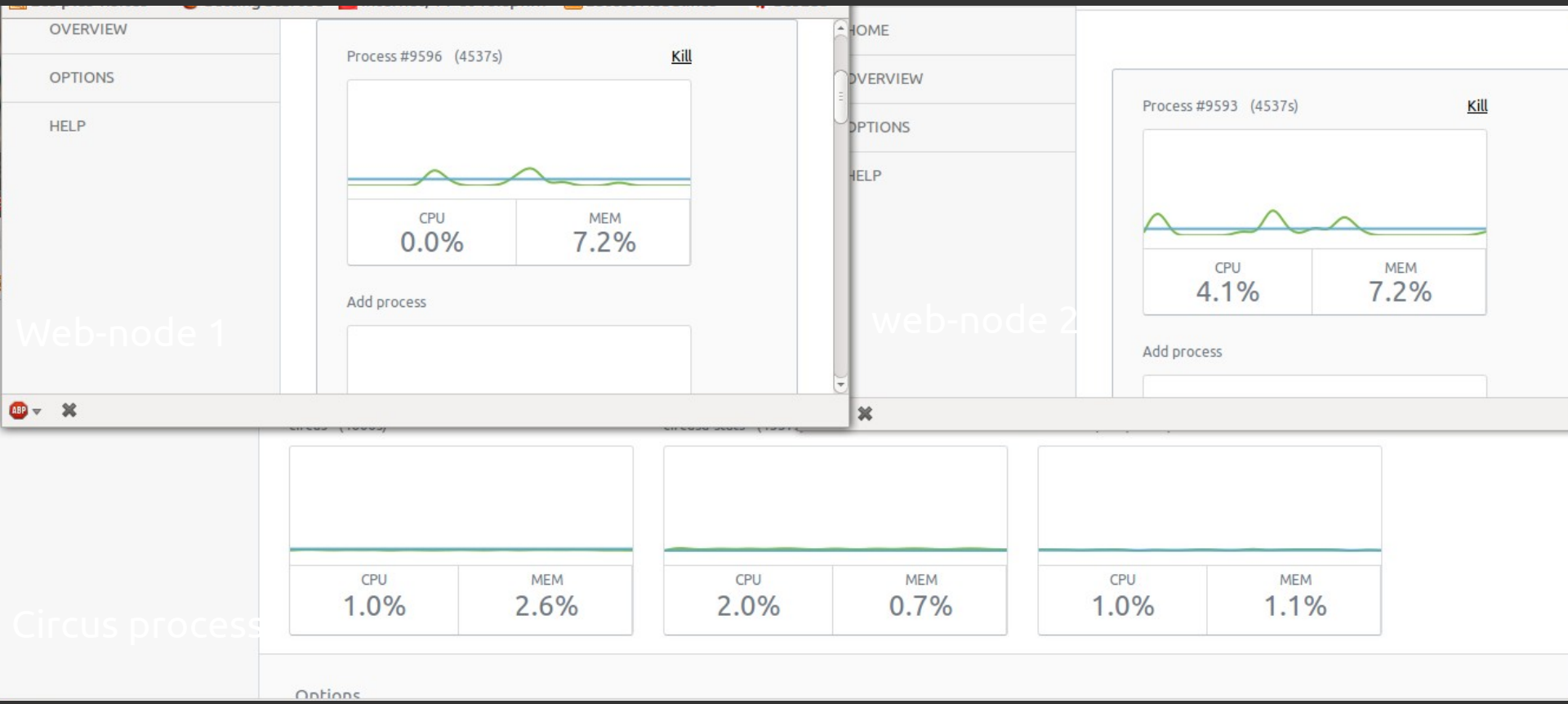
Circus Daemons

CPU Memory

circus (4899s)		circusd-stats (4836s)		circushttpd (4836s)	
CPU	MEM	CPU	MEM	CPU	MEM
1.0%	2.6%	3.1%	0.7%	0.0%	1.1%

Options

ZeroMQ endpoint used to connect clients like *circusctl* or *circushttpd*	tcp://127.0.0.1:5555
ZeroMQ endpoint used for pub/sub	tcp://127.0.0.1:5556
ZeroMQ endpoint used for the pub/sub stats	tcp://127.0.0.1:5557
Delay in seconds for the ZMQ polling.	1.0


OVERVIEW
OPTIONS
HELP
Process #9596 (4537s)
Kill
CPU
0.0%
MEM
7.2%
Add process
Web-node 1
HOME
OVERVIEW
OPTIONS
HELP
Process #9593 (4537s)
Kill
CPU
4.1%
MEM
7.2%
Add process
web-node 2
CPU
1.0%
MEM
2.6%
CPU
2.0%
MEM
0.7%
CPU
1.0%
MEM
1.1%
Circus process
Options