

Coalition : a simple and useful tool to distribute R-works on a set of computers

Marie-Pierre Etienne¹, Cyril Corvazier² and Benjamin Legros²

1. AgroParisTech - INRA
2. Mercenaries Engineering



User! 2009 Conference

marie.etienne@agroparistech.fr

Coalition is a task scheduler

Coalition - Mozilla Firefox

http://gauss.agroparistech.fr:8080/

Most Visited Smart Bookmarks Google Calendar

Coalition

A small but beautiful job manager by Mercenaries Engineering - GPL licence

Jobs Workers Refresh

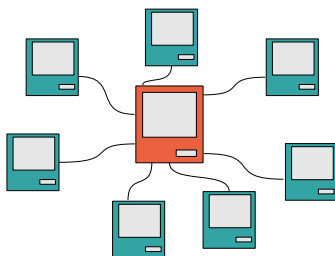
Clear All jobs Remove Selection Reset Selection

1000 Set Selection Priority

Order	ID	Title	User	State	Priority	Affinity	Worker	Duration	Try	Command	Dir	Dependencies	Tools
1	126	job	lebarbie	FINISHED	1000		aristote64_1	3 d 9 h 39 m 23 s	1/10	Rscript LauchFitSegLM3.R	Simu\MVA\PGM		Log Remove Reset
2	525	New job	metienne	FINISHED	1000		galileo64_3	1 d 0 h 27 m 15 s	1/10	Rscript BootStratStudy.R 1000 0	data\Recherche\PoisComp		Log Remove Reset
3	563	essai variance	lang	ERROR	999		kuiper64_1	2 s	1/1	Rscript variance.r	.		Log Remove Reset
4	670	New job	metienne	FINISHED	990		galileo64_4	18 h 12 m 48 s	1/10	Rscript jobSunStar.R 50 20 30 150	data\Recherche\PoisComp		Log Remove Reset
5	671	New job	metienne	WORKING	990		kuiper64_4	19 h 14 m 3 s	1/10	Rscript jobSunStar.R 50 20 40 150	data\Recherche\PoisComp		Log Remove Reset
6	675	New job	metienne	FINISHED	990		kuiper64_1	6 h 15 m 52 s	1/10	Rscript jobSunStar.R 50 20 5 200	data\Recherche\PoisComp		Log Remove Reset
7	676	New job	metienne	FINISHED	990		aristote64	38 h 31 m 47 s	1/10	Rscript jobSunStar.R 50 20 10 200	data\Recherche\PoisComp		Log Remove Reset
8	677	New job	metienne	FINISHED	990		aristote64_1	10 h 57 m 52 s	1/10	Rscript jobSunStar.R 50 20 15 200	data\Recherche\PoisComp		Log Remove Reset
9	678	New job	metienne	FINISHED	990		aristote64_3	13 h 24 m 9 s	1/10	Rscript jobSunStar.R 50 20 20 200	data\Recherche\PoisComp		Log Remove Reset
10	679	New job	metienne	WORKING	990		galileo64_3	13 h 25 m 33 s	1/10	Rscript jobSunStar.R 50 20 25 200	data\Recherche\PoisComp		Log Remove Reset
11	680	New job	metienne	WORKING	990		aristote64_2	212 h 38 m 57 s	1/10	Rscript jobSunStar.R 50 20 30 200	data\Recherche\PoisComp		Log Remove Reset
12	681	New job	metienne	WORKING	990		galileo64_2	12 h 34 m 36 s	1/10	Rscript jobSunStar.R 50 20 40 200	data\Recherche\PoisComp		Log Remove Reset
13	683	New job	metienne	WORKING	990		kuiper64_1	11 h 49 m 0 s	1/10	Rscript SaintLaurentSimu.R 500 4000	data\Recherche\PoisComp		Log Remove Reset
14	684	New job	metienne	WORKING	990		kuiper64_2	9 h 39 m 59 s	1/10	Rscript SaintLaurentSimu2.R 500 4000	data\Recherche\PoisComp		Log Remove Reset
15	685	New job	metienne	WORKING	990		kuiper64_3	5 h 51 m 48 s	1/10	Rscript jobSunStar.R 50 20 5 250	data\Recherche\PoisComp		Log Remove Reset
16	687	New job	metienne	WORKING	990		aristote64_3	5 h 51 m 40 s	1/10	Rscript jobSunStar.R 50 20 10 250	data\Recherche\PoisComp		Log Remove Reset
17	688	New job	metienne	WORKING	990		galileo64_1	5 h 27 m 2 s	1/10	Rscript jobSunStar.R 50 20 15 250	data\Recherche\PoisComp		Log Remove Reset
18	689	New job	metienne	WORKING	990		aristote64_1	5 h 23 m 45 s	1/10	Rscript jobSunStar.R 50 20 20 250	data\Recherche\PoisComp		Log Remove Reset
19	690	New job	metienne	WORKING	990		galileo64_4	2 h 42 m 13 s	1/10	Rscript jobSunStar.R 50 20 25 250	data\Recherche\PoisComp		Log Remove Reset
20	691	New job	metienne	WORKING	990		aristote64_4	1 h 57 m 20 s	1/10	Rscript jobSunStar.R 50 20 30 250	data\Recherche\PoisComp		Log Remove Reset
21	692	New job	metienne	WAITING	990		0	0 s	0/10	Rscript jobSunStar.R 50 20 40 250	data\Recherche\PoisComp		Log Remove Reset
22	693	New job	metienne	WAITING	990		0	0 s	0/10	Rscript jobSunStar.R 50 20 5 300	data\Recherche\PoisComp		Log Remove Reset

Coalition principle

- One server schedules tasks running with `server.py` script
- Workers execute tasks running with `worker.py` script



- Coalition is available at <http://code.google.com/coalition/> under GNU General Public License v3

How to start with Coalition

1 Running the Server

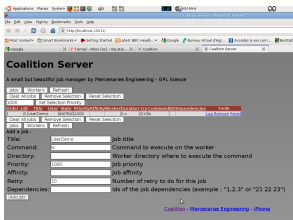
```
python server.py
```

2 Running the Worker

```
python worker.py
```

How to use Coalition ?

1 Using Web interface



2 Using a Python script control.py

```
python control.py -c "ls" -t "UserDemo2"
http://localhost:19211 add
```

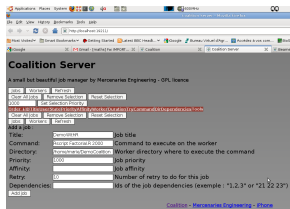
How to use Coalition with R?

Use of Rscript to run R in command line.

Factorial.R is located
/home/metienne/DemoCoalition
and contains :

```
## the wrong way to compute
##factorial of a given argument
args <- commandArgs(TRUE)
m.max <- type.convert(args[1])
file.out <- paste("factorial",
                 m.max, ".txt", sep="")
factorial <- numeric(m.max)
for( i in 1:m.max)
{
  prov <- 1
  for(j in 1:i)
  {
    prov <- prov*j
  }
  factorial[i] <- prov
}
# write results in file
write.table(factorial,file.out)
```

1 Using Web interface



2 Scripting with control.py

```
python control.py
-c "Rscript Factorial.R 1000"
-d "/home/metienne/DemoCoalition"
http://localhost:19211 add
```

Other basic functions

Every function may be controlled using either web interface or `control.py` script.

- Show logs
- Remove a job/a selection of jobs
- Reset a job/a selection of jobs
- Control workers

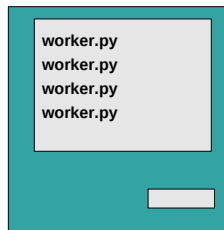
Priority, Dependency and Affinity system

- **Priority.** To give some priority to pressing jobs, change the priority level : Jobs are submitted according to their priority level.
- **Dependency.** If your job, say number 10, needs the results of another one say 8, use dependency option. Since submitting the job, precise the job ID of the required one.
- **Affinity.** You can tag workers with affinities. A job requiring specific affinities will be run only on a worker meeting all the specified affinities. It is a way to manage R-packages availability on a pool of computers.

Multicore management

To fully exploit the multicore capacities of the processor, one simply runs a worker per core available on the computer. Be sure to have enough RAM for all the processes.

Quad Core



Coalition deals also with

- LDAP authentication
- Windows and GNU/Linux support
- iPhone support ...

And aims to deal with

- Kinship notion and progress status
- Sampling next task according to priority level

To conclude

- Coalition is a solution to share computational resources.
- If you can divide your work in independent tasks, Coalition allows an optimal usage of your computational resources.
- Using OS scheduler, worker may be launched during inactivity time.
- And last but not least, Coalition is very simple to deploy.

Thanks to Hamid Aichoune, our IT, for his advices and his beta testing work.