



Revisiting virtual machine consolidation to save resources and energy in heterogeneous production cloud infrastructures

Simon Lambert, Vladimir Ostapenco, Laurent Lefèvre Eddy Caron, Anne Cécile Orgerie, Benjamin Fichel, Rémi Grivel 26/03/2025











Context

Main problems :

- Cloud Service Providers (CSP) size their infrastructure to meet peak demand.
- CSPs are reluctant to use resource over-subscription.
- → Results in oversized infrastructures, waste of resources, energy etc.

Different works on elasticity, shutdown techniques or consolidation to reduce these resource and electricity consumption.

Not used in production most of the time for economical or technical reasons.









Methodology

Versatile consolidation methodology to facilitate production implementation.









Infrastructures

Tested on virtualization clusters of two cloud providers :

OVHcloud

1 cluster (L1) :

985 Physical Machines (PMs) 7376 Virtual Machines (VMs)



Innia

2 clusters (S1 & S2) :

6 PMs (S1) ; 16 PMs (S2) 400 VMs (S1) ; 1000 VMs (S2)





DELYON



Methodology steps



Methodology step	Small scale	Large scale	
Detect consolidation potential	Initial infrastructure allocation ratios (CPU, memory, storage etc) and usage study		
Low Resource Usage (LRU) VM detection	LRU VM algorithm on CPU and network	LRU VM algorithm on CPU, network and disk	
Consolidation strategies	1 strategy : all VMs with oversubscription	4 strategies based on VM scope and use of oversubscription	
VM placement	First Fit Decreasing heuristic to reduce number of PMs	Optimal Bin Packing Solver to reduce number of PMs and migrations	
Allocation ratios study	Based on manual experimentation RAM (1/2/3)	Based on hypervisor documentation CPU (1/2/4/8) and RAM (1/1.5/2)	
Simulation process	Python simulator based on clusters traces	SimGrid based on cluster traces	





Une marque de Ciril GROUP

cir

LRU VMs algorithm



LRU VMs detection algorithm :

```
is_used(vm, t, cpuRate, netRate, diskRate):
if cpu_usage(vm,t) <= cpuRate and net_usage(vm, t) <= netRate [and disk_usage (vm, t) <= diskrate]:
    return 0
else:
    return 1</pre>
```

Studied thresholds :

- CPU : 1%, 2%, 5% and 10%
- I/O (Network and disk) :
 - Small scale : 1Kb/s, 200Kb/s and 2Mb/s
 - Large scale : 200KB/s, 2MB/s and no threshold







EESULTS Small scale – Consolidation potential and LRU VMs

Unused hosts	CPU usage (%)	RAM usage (%)	
0	30 - 60%	55 – 70%	

Small scale clusters aggregated statistics



S1 cluster - LRU VMs percentage



S2 cluster - LRU VMs percentage

OVHcloud



HE LYES







Small scale – Consolidation potential



S1 cluster – VM usage distribution over one week

Mean usage Quantile Range 70 -60 50 Usage (%) 40 30 20 10 0 20 15 Ó 5 10 Hour of the Day

S1 cluster – Temporal VM usage analysis

Results with a CPU threshold of 1% and I/O threshold of 1kbps







Small scale – Energy savings



S1 cluster – Energy savings

Considered VMs : LRU-80 ; Oversubscription ratio : 3



(naío-

S2 cluster – Energy savings





RAFACERSILIE RAFACERSILIE







Small scale – Energy savings (LRU 80 VMs)



CPU threshold : 1% I/O threshold : 1kbps Considered VMs : LRU-80 Oversubscription ratio : 3

\rightarrow 22.05% power consumption reduction

S1 cluster – Power consumption with and without consolidation







Large scale – Freed hosts



L1 cluster - Number of freed hosts for the LRU VMs migration strategy and the LRU ratio of 99%



L1 cluster - Number of freed hosts for the general VM migration with oversubscription strategy. (CPU 10% and no I/O threshold)

V OVHcloud

SynAApS









Conclusion and perspectives

Versatile consolidation methodology :

- Multiple methodology steps tested on heterogeneous infrastructures
- Enable resources and energy consumption reduction in both cases

Perspectives :

- Assess technical constraints for each infrastructure typology
- Evaluate multiple environmental impact reductions
- Include users to enhance consolidation results











Thank you.

@mail : simon.lambert@ens-lyon.fr







14 Green Days 2025



Large scale – Consolidation potential and LRU VMs

Hosts	VMs	Unused hosts	Average host CPU usage
985	7376	20	9.3%

Large-scale Cluster L1 statistics

	CPUs		RAM		sk
Total	Allocated	Total	Allocated	Total	Allocated
51.1k	66.5%	145.8 TB	90.7%	1234.7 TB	64.2%

Resource allocation statistics for used hosts



200KB/s

2MB/s

Network and Disk thresholds

L1 cluster - LRU VMs percentage

CPU usage thresholds (%)

None

200KB/S

5

2MB/S

Network and Disk thresholds

10

None

Inría















Large scale – Energy savings



L1 Cluster - Percentage of energy saved.

Strategy A - LRU VMs migration strategy for the most relaxed thresholds. Strategy B - General VM migration strategy without oversubscription. Strategy C - General VM migration with oversubscription and allocation ratios of 2 for both CPU and RAM





