Scheduling Challenges in a Shared Private Cloud Infrastructure

Dalibor Klusáček[1] and Boris Parák[1] and Lukáš Hejtmánek[2]
[1]CESNET a.l.e., Czech Republic — [2]Institute of Computer Science, Masaryk University, Czech Republic
{klusacek,parak}@cesnet.cz, xhejtman@ics.muni.cz

1. Motivation

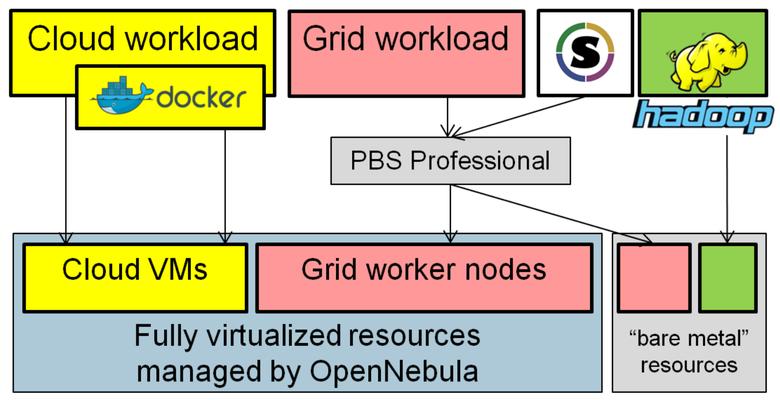
This work describes **scheduling challenges** in a shared private cloud infrastructure. Using real-life data we analyze following issues observed in the system:

- impact of "free of charge" computing
- reclaiming of inactive resources
- cloud (under)utilization
- fairness-related issues in cloud

2. Shared Cloud-based Infrastructure

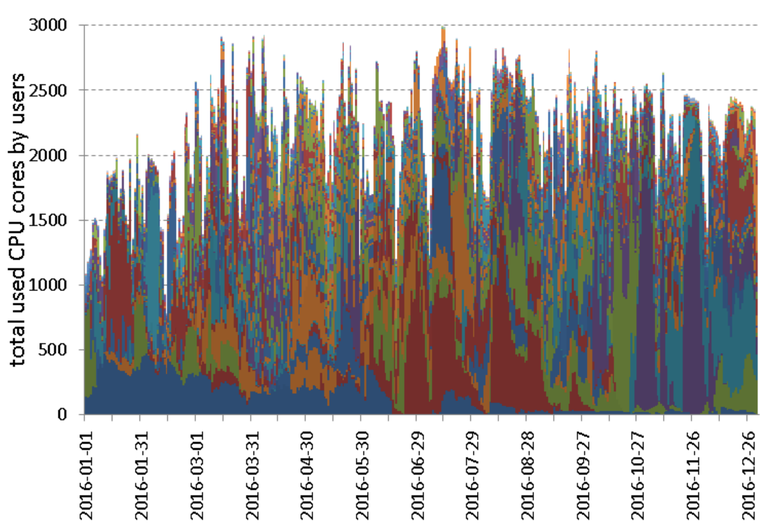
***MetaCentrum* infrastructure**:

- infrastructure is mostly virtualized
- currently using OpenNebula platform
- delivering flexible IaaS
- VMs may host grid worker nodes
- PBS-Pro uses grid worker nodes
- load-balancing done "by hand"



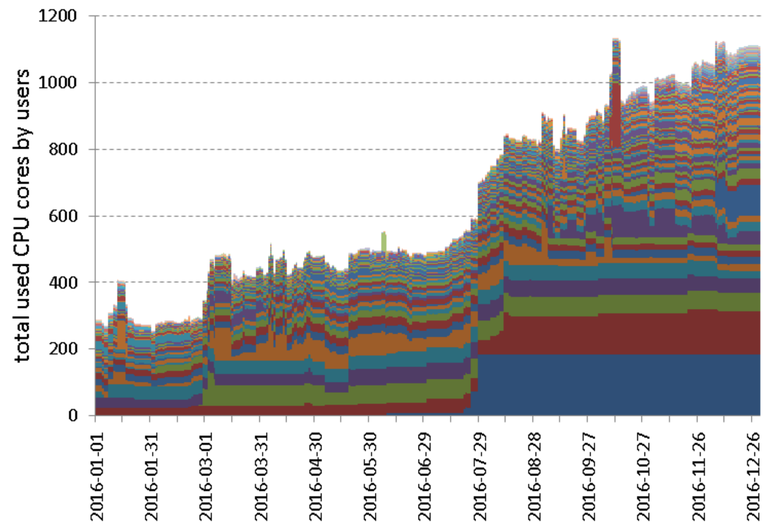
3. Workload Comparison Between Grid and Cloud Partition

CPU core utilization by users in the grid



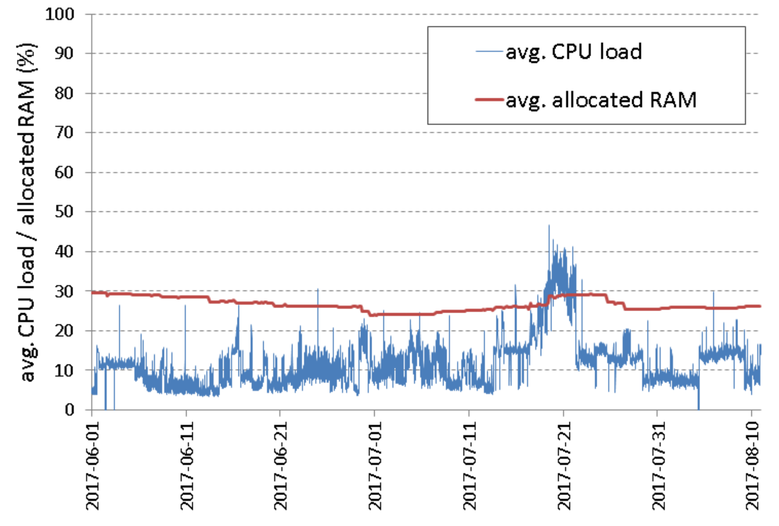
Grid usage is controlled by fair-share. Each job has a maximum walltime limit.

CPU core utilization by users in the cloud



VMs run much longer than grid jobs. No fair-share analogy (or $ billing) is used.

CPU load and RAM allocation in the cloud



Actual load of running VMs is very low causing resource wasting.

4. Proposed Scheduling Approaches and Further Opportunities

Newly deployed solutions in the cloud

- **limited VM lifetime** (3 months)
- explicit prolongation request needed
- automatic killing of "zombie" VMs

Considered ways to improve utilization and fairness in the cloud partition

- advanced VM scheduler [1]
- "scavenge computing" (short grid jobs) using CPU cycles of idle VMs
- fair-share-inspired VM prioritization
- used for adjusting dynamic VM migrations and overbooking ratios

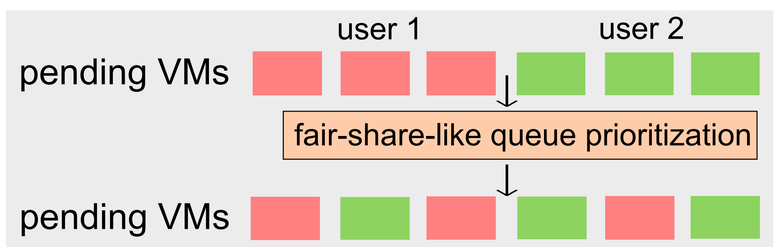
Observed opportunities

- **CPU load of running VMs is low**
- same applies for the allocated RAM
- idle nodes can be used otherwise
- nights/weekends are safe to run, e.g., batch grid jobs in the background
- **91% of jobs request less than 12 hours**
- such jobs represent 5% of the total CPU usage in CERIT-SC system

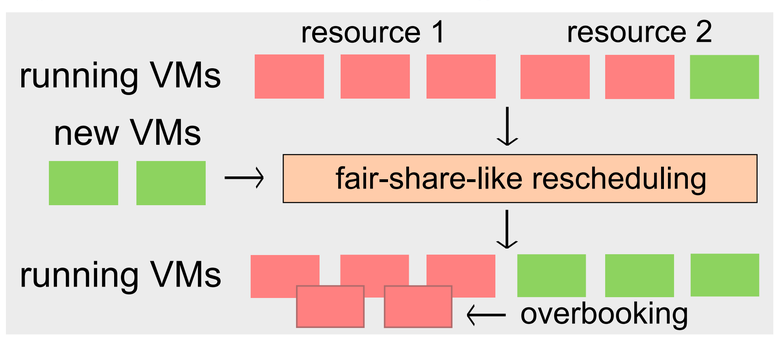
"Scavenge-like" grid computing

- using "grid worker" VMs on nodes
- and special queue in the PBS-Pro

Dynamic VM prioritization



Dynamic VM rescheduling



5. Conclusion and Future Work

- virtualization enables resource sharing between cloud and grid [3]
- current setup causes problems, e.g., low resource utilization [2]
- new policies are deployed and tested
- further optimization and development will be needed in the future
- e.g., user and VM prioritization to reflect fairness and various SLOs
- dynamic overbooking

Acknowledgments

We kindly acknowledge the support provided by the MetaCentrum under the program LM2015042, the CERIT Scientific Cloud under the program LM2015085 and the project Reg. No. CZ.02.1.01/0.0/0.0/16_013/0001797 co-funded by the Ministry of Education, Youth and Sports of the Czech Republic.

References

[1] G. Podolníková, B. Parák and D. Klusáček. Extensible and Modular Cloud Scheduler for OpenNebula. In *Cracow Grid Workshop*, AGH, 2015.

[2] D. Klusáček and B. Parák . Analysis of Mixed Workloads from Shared Cloud Infrastructure. In *Job Scheduling Strategies for Parallel Processing*, 2017.

[3] D. Klusáček and G. Podolníková. Scheduling Hybrid Workloads in Shared Cloud Infrastructures. In *Cracow Grid Workshop*, AGH, 2016.