Experience in using public Cloud for Belle II experiment within HNSciCloud Project

S. Pardi[1] D. Michelino[1,2]
[1]INFN-Napoli Unit – Italy
[2]GARR – Italy

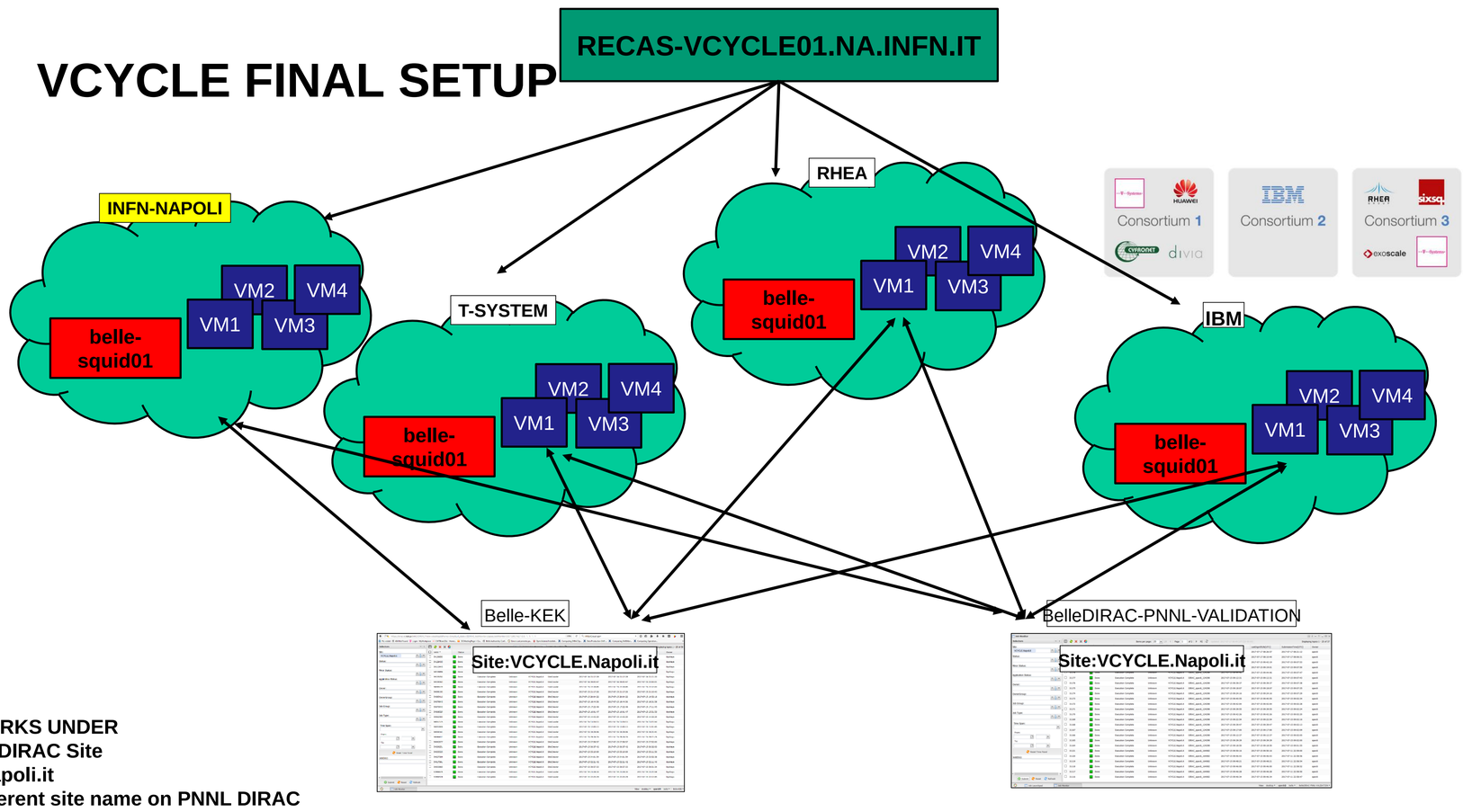
INTRODUCTION

The HNSciCloud is a European pre-commercial procurement (PCP) project co-funded by the European Commission Horizon 2020 which aims at creates a competitive marketplace of innovative cloud services for scientific user communities.

In this poster we show the activities done to support the Belle II use-case. In particular the final goal is to run standard Monte Carlo production jobs and user-analysis on the Public Clouds selected by the project, without bottleneck.

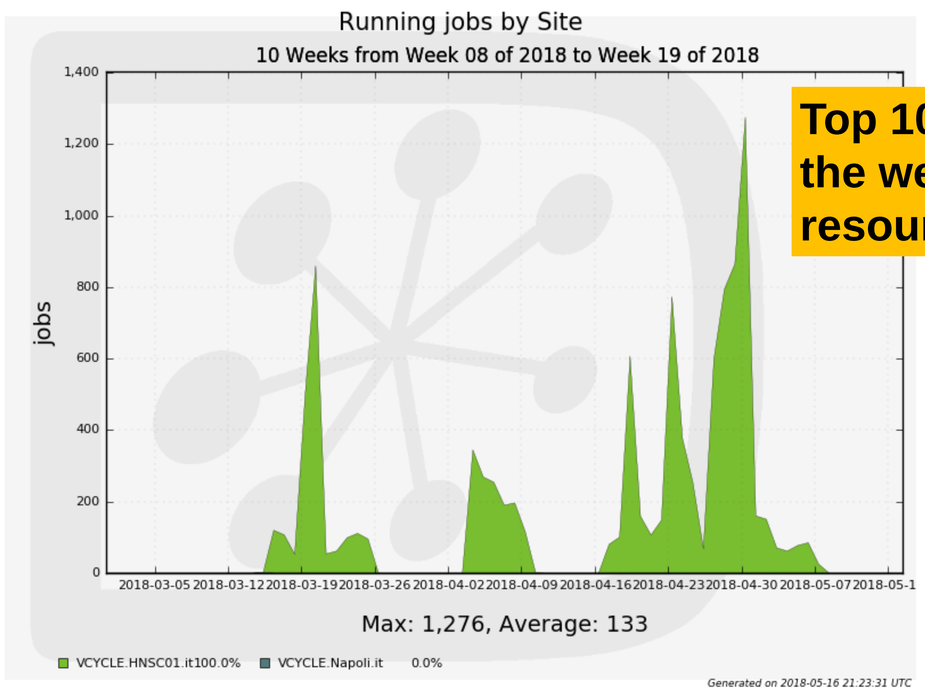
IMPLEMENTED SOLUTION

For each cloud, we implemented a cache service based on squid registered as well in a local DNS (for the OTC based clouds). The orchestrator has been centralized on a single server located in the INFN-Napoli infrastructure (ReCaS-Napoli datacenter).
As regarding he network setup, in the case of IBM Cloud all the VMs have been configured with public IP address, while in case of T-system and RHEA solution, we used private IP and a SNAT server, created ad HOC with 10Gbps connection to the WAN.
Cloud systems has been integrated in Belle II frameworks based on DIRAC by using VCYCLE as Virtual Machine Life Cycle Manager.
Finally we implemented a script that optimize the number of VMs to create of a specific flavor, in order to use the available resources up to quotas limits.

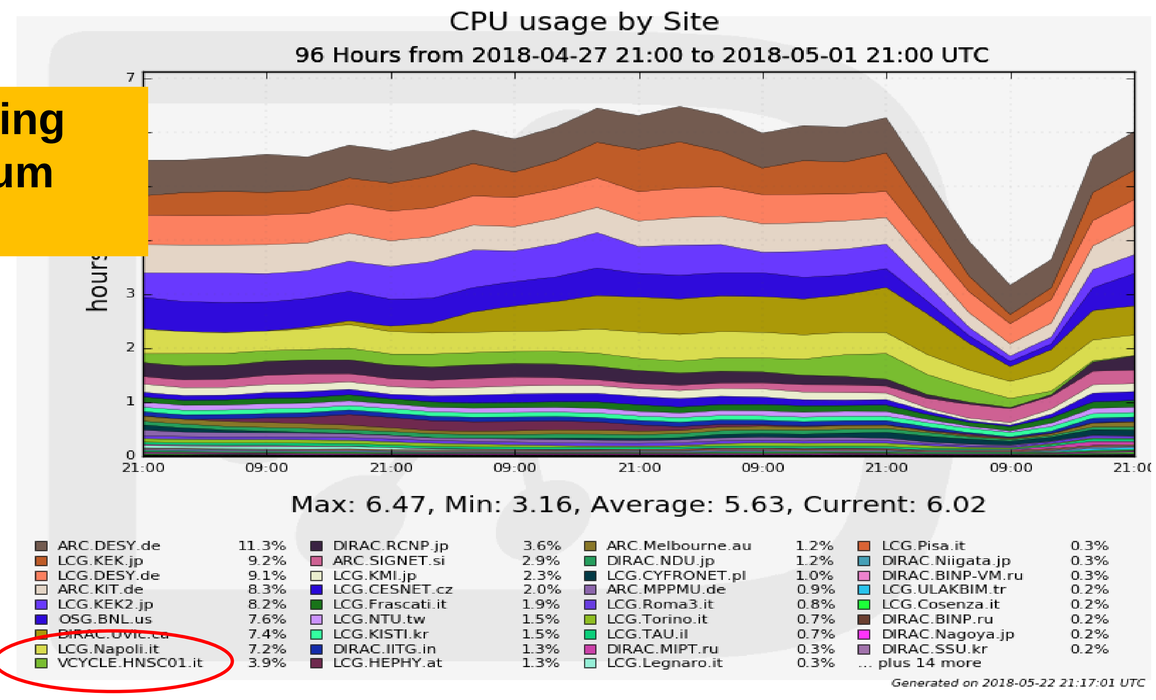


Used Flavors:
- 1 CPU, 4GB RAM, 40GB disk
- 2 CPU, 8GB RAM, 60GB disk
- 4 CPU, 16GB RAM, 100GB disk
- 8 CPU, 32GB RAM, 180 GB disk

CLOUD WORKS UNDER THE SAME DIRAC Site VCYCLE.Napoli.it As with different site name on PNNL DIRAC VCYCLE.HNSC01.it, VCYCLE.HNSC02.it, VCYCLE.HNSC03.it



Top 10 reached during the week of maximum resource allocation



CONCLUSION

After a tuning phase, we have been able to take advantage from the resources provided by Public Cloud providers using standard Belle II framework tools and VCYCLE as Virtual Machine Management System.
During the scalability phase, we demonstrated to be able to use all the available resources on the cloud, growing up to the quota limits.
The lesson learned by this experience is that, even if plug-and-play approach is possible, for a proper usage of Public Cloud a certain amount of time must be take in count to configure the environment before to run properly and efficiently user-specific applications. Preliminary steps may be different in function of the specific platform ran by the Cloud Provider, among them the network setup represent one of the most laborious point. The second aspect to take in account is the balance of different quotas (ie #CPU, RAM, number of machine, number of disks) that represents one of the key points to run applications in optimal condition over a pay-per-use infrastructure.