

Šustr, Z.; Sitera, J.; Mulač, M.; Ruda, M.; Antoš, D.; Hejtmánek, L.; Holub, P.; Salvét, Z.; Matyska, L.

CESNET, <http://meta.cesnet.cz>

Introduction

In recent years, progress in MetaCentrum has relied heavily on the concepts and technologies of virtualization. A major portion of MetaCentrum computation resources is currently virtualized, allowing the physical infrastructure to be shared by EGEE nodes. This arrangement is ideal for the gradual introduction of EGI, as it will allow us to keep running legacy gLite-based infrastructure for the benefit of users who cannot switch to UMD-based services at once.

The virtual nature of MetaCentrum resources is mostly transparent to end users, though, due to their integration with the resource management system. We are currently introducing a new public “virtual cluster” service, however, making the virtualized infrastructure accessible as an end user service.

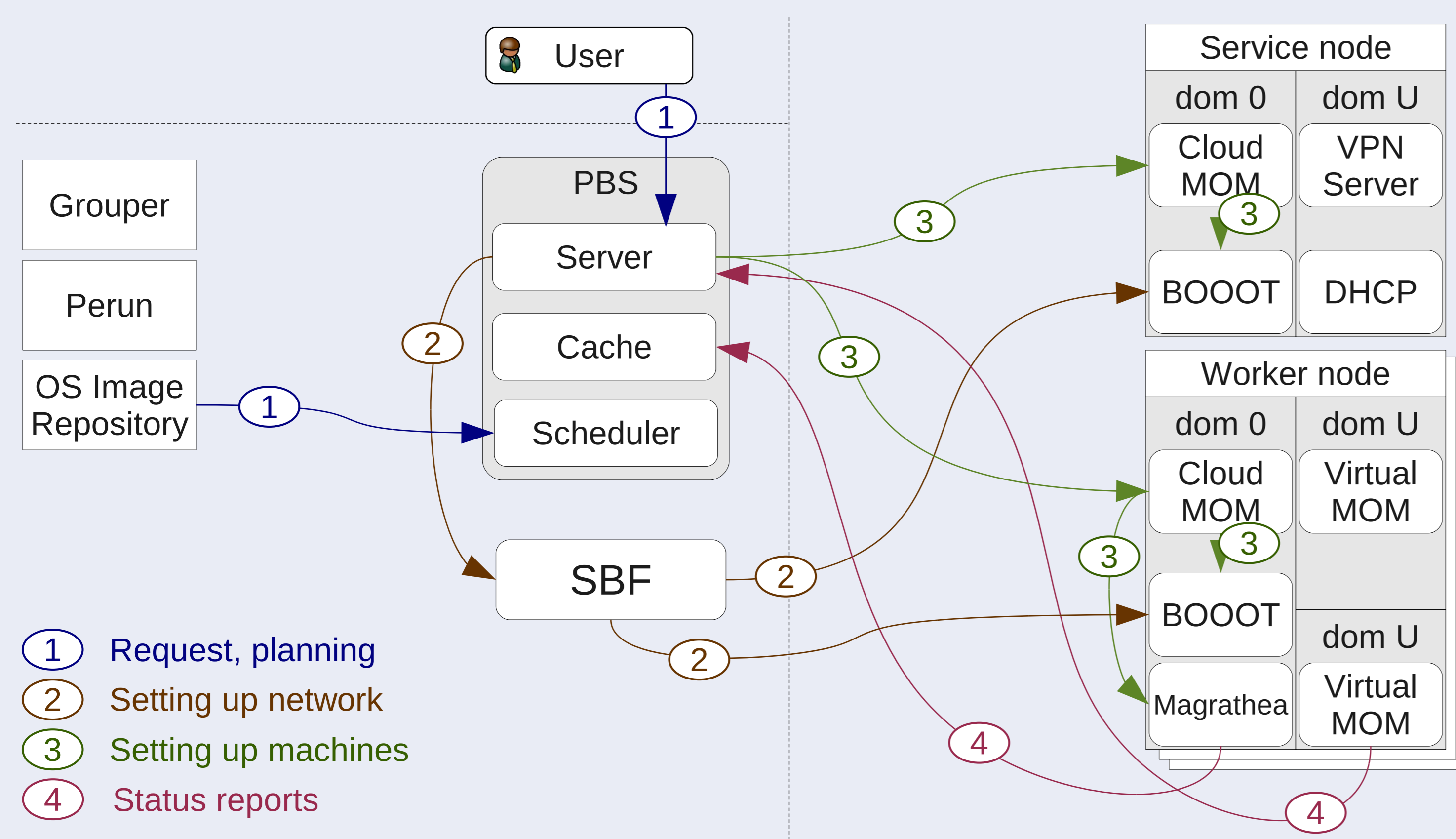


The MetaCentrum project, an activity of the CESNET association, covers a majority of activities concerning Grids, cluster and high performance computing across the Czech Republic. MetaCentrum operates and manages a distributed computing infrastructure consisting of computing and storage resources owned by CESNET and other co-operating academic institutions within the Czech Republic. MetaCentrum is responsible for building the National Grid and integrating it into relevant international infrastructures, especially those funded by the European Union. It is actively participating in many international Grid projects such as EGEE, EGI_DS or EUAsiaGRID.

Architecture

Components and Interactions

The overall MetaCentrum architecture consists of four major components as illustrated below:

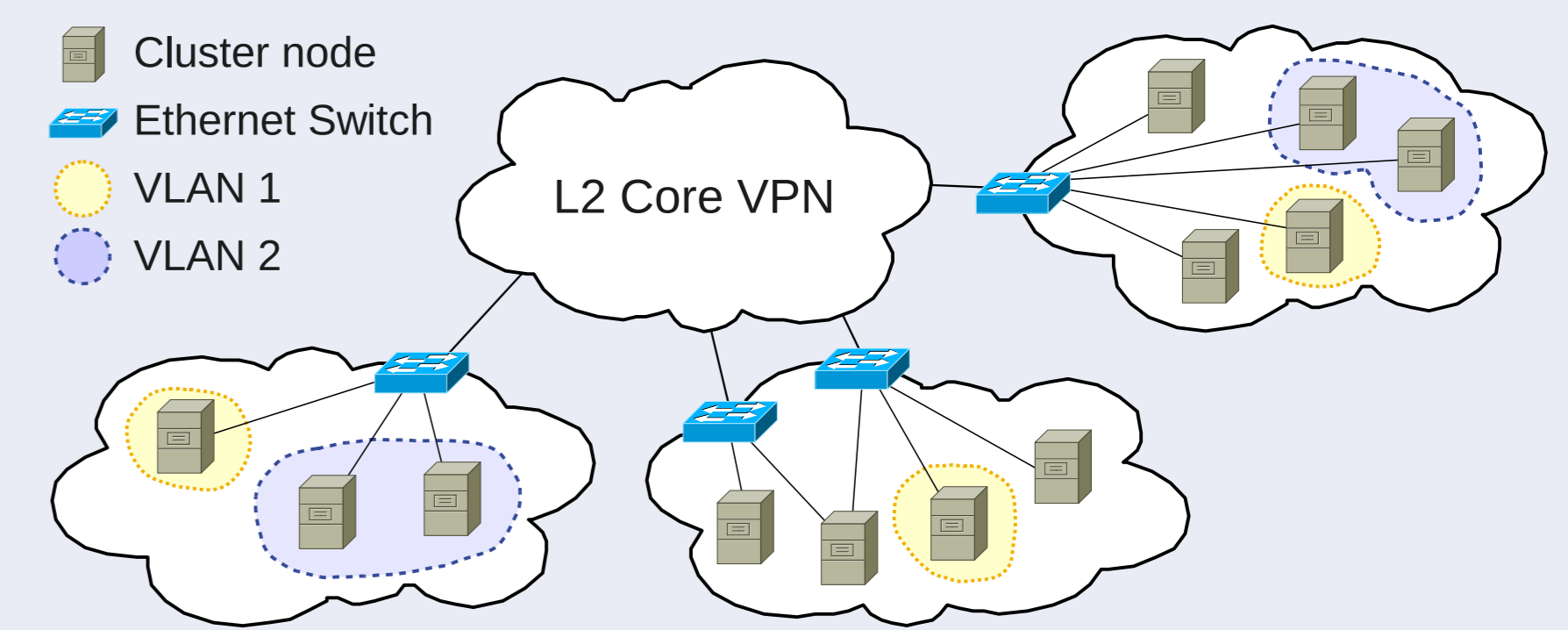


- ① Request, planning
- ② Setting up network
- ③ Setting up machines
- ④ Status reports

- Virtual appliances management is responsible for controlling the overall virtual cluster life cycle. It relies on a modified batch system (PBS) coupled with Magrathea (see below), managing virtual clusters in the same way as jobs.
- Virtual network management (SBF) is an executive unit providing a “network” resource (L2 VLAN).
- Worker node – physical resource hosting virtual cluster machines. The “BOOOT” Component sets up virtual machines using appropriate OS images and configurations (account information received from Perun information service) and connects them to their respective virtual networks.
- Service node – there may be one service node in every virtual cluster, running dedicated infrastructure services such as a VPN server, which makes the cluster accessible, or a DHCP server. The service node may also serve as a gateway to other networks/services.

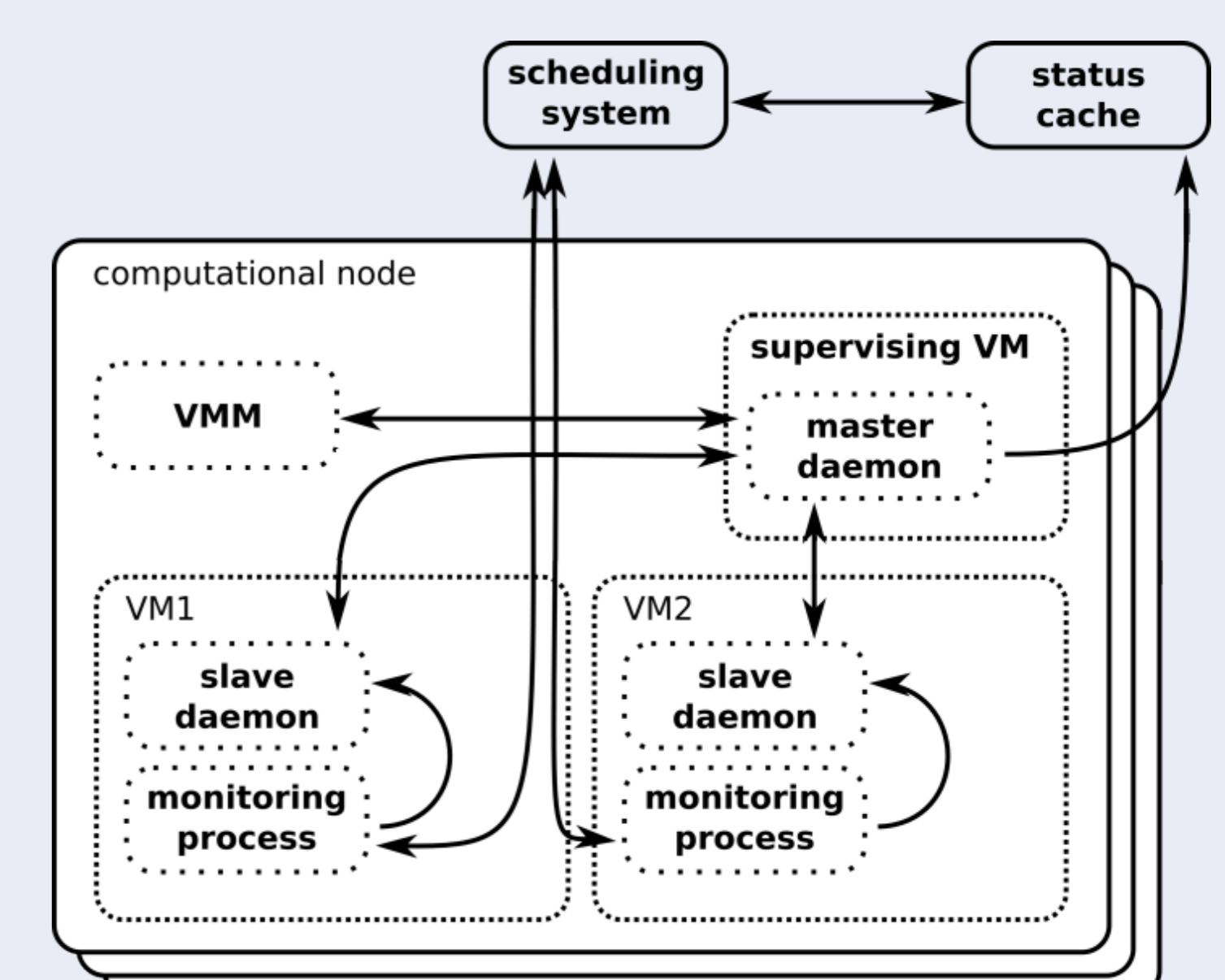
Networking

In our virtual cluster service, the network is just another resource made available to users, managed and allocated by a resource manager. Traditionally, networks are seen as “fixed grid resources,” but this is not sufficient for virtual clusters. Virtual clusters are mapped dynamically to individual resources provided by the physical infrastructure, the network being no exception. That is why we have designed VirtCloud [2], a system for setting up dynamic virtual clusters over a nation-wide network, supporting cluster encapsulation and publishing. The VirtCloud system requires, by design, no run-time reconfiguration of the core network. Network resources (nation-wide L2 VLAN) can be created and assigned with no significant overhead. The architecture has been prototyped in the Grid environment of the MetaCentrum project using the CESNET2 backbone network [1].



Magrathea

Magrathea [3] provides an interface between the batch scheduling system and virtual machine monitors. It allows jobs to be submitted in a controlled manner into virtual machines running on cluster nodes (with multiple virtual machines running on a single cluster node). Magrathea can also be seen as a scheduler running on each cluster node, scheduling virtual machines according to jobs submitted by batch scheduling systems. Current implementation supports PBS Pro (modified slightly to support all of Magrathea’s features) and Xen as well as VServer virtual machine monitors.



References

- [1] Novák, V.; Šmrha, P.; Verich, J.: *Deployment of CESNET2+ E2E Services*, CESNET, z. s. p. o., Tech. Rep. 18/2007, December 2007.
- [2] Antoš, D. et al.: *VirtCloud: Virtualizing Network for Grid Environments—First Experiences*; The IEEE 23rd International Conference on Advanced Information Networking and Applications, 29 May 2009, Bradford, UK.
- [3] Ruda M. et al.: *Scheduling Virtual Grids: the Magrathea System*, Second International Workshop on Virtualization Technology in Distributed Computing, 2007, Reno, USA.

Benefits

- Highly flexible resource sharing for multiple middleware environments
- Benefiting from the very advanced environment of the CESNET2 NREN

Typical Use Cases

- **Safe environment.** Being protected from the outside world, users may continue running their virtual systems even with weak or outdated safety measures.
- **Exotic configurations.** Users do not need to adjust their jobs to run in the grid environment. They may just supply their own OS image and run their tasks.
- **Easy adoption.** Once an existing cluster is virtualized, its users can start utilizing superior grid resources easily, requiring no extra training.
- **Private networking.** Virtual machines may be connected to workgroup VLANs, making them almost unrecognizable from hardware located physically on users’ premises. The virtual nature of such machines is totally transparent to users.



The EGEE project is building a Grid infrastructure for the scientific community. Grids are networks of computers spread across many sites but able to act together to provide a range of large scale facilities, from incredible processing power and mass storage to a platform for international collaboration.