

R&D at CNAF

The WNoDeS 2 Framework

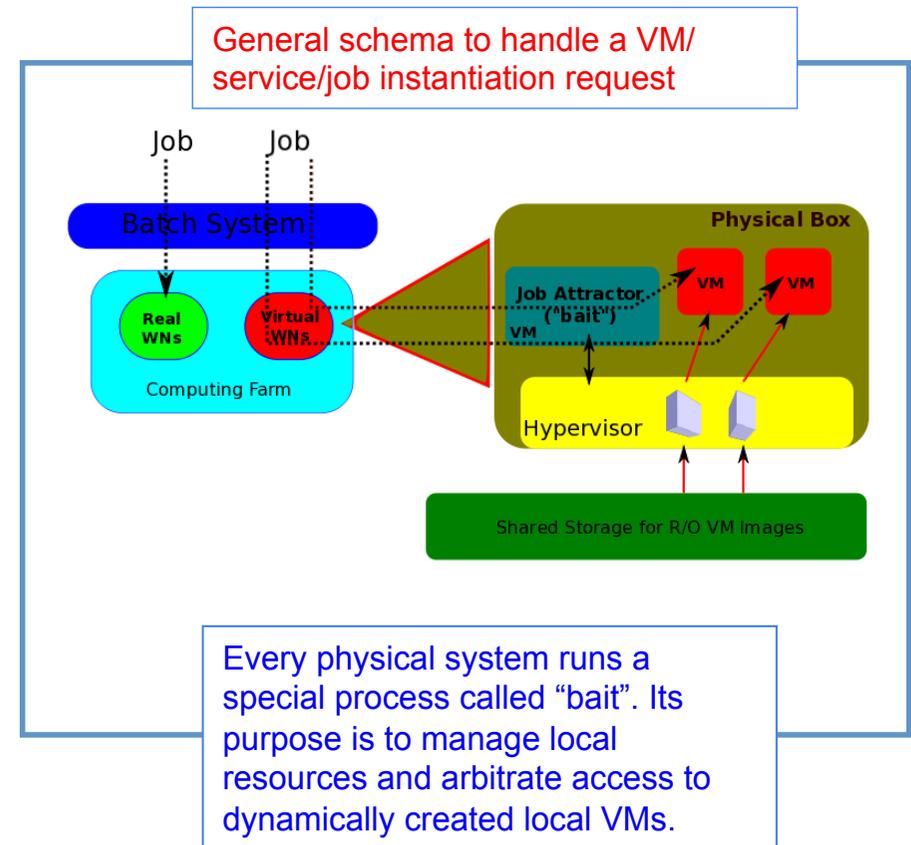
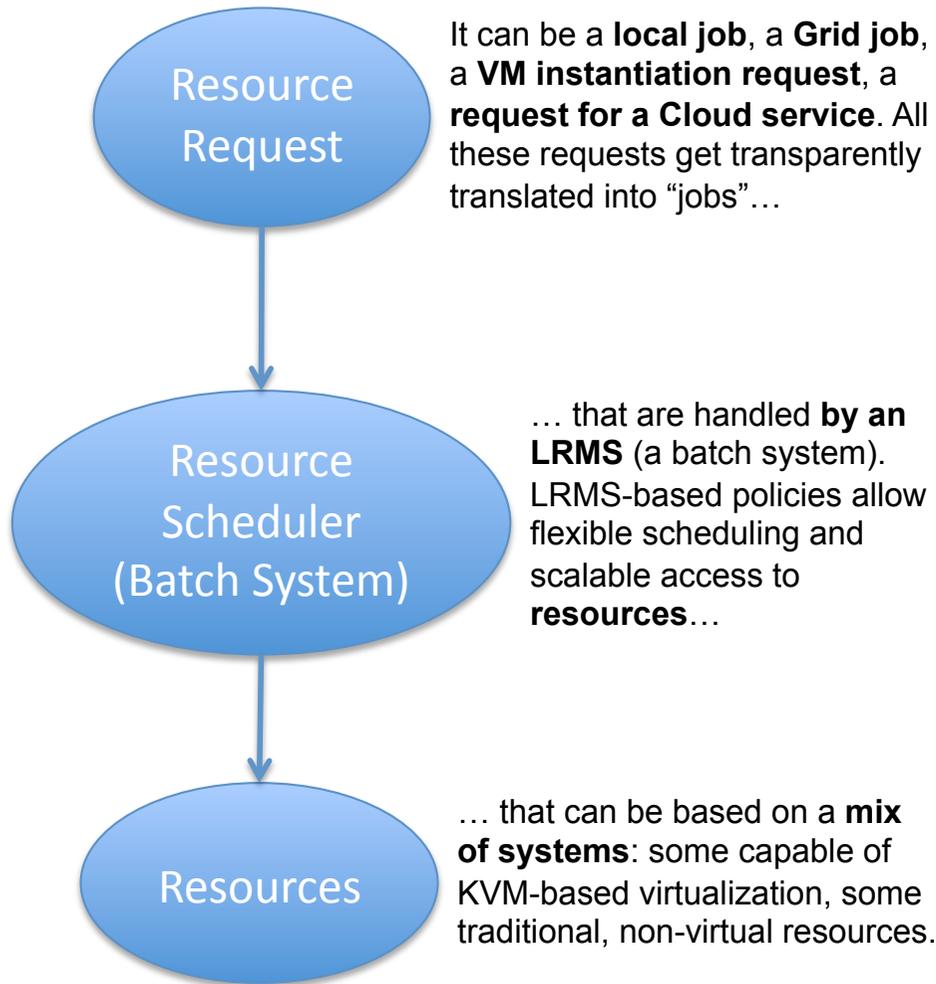
Davide Salomoni, INFN

Fermilab, May 18, 2012

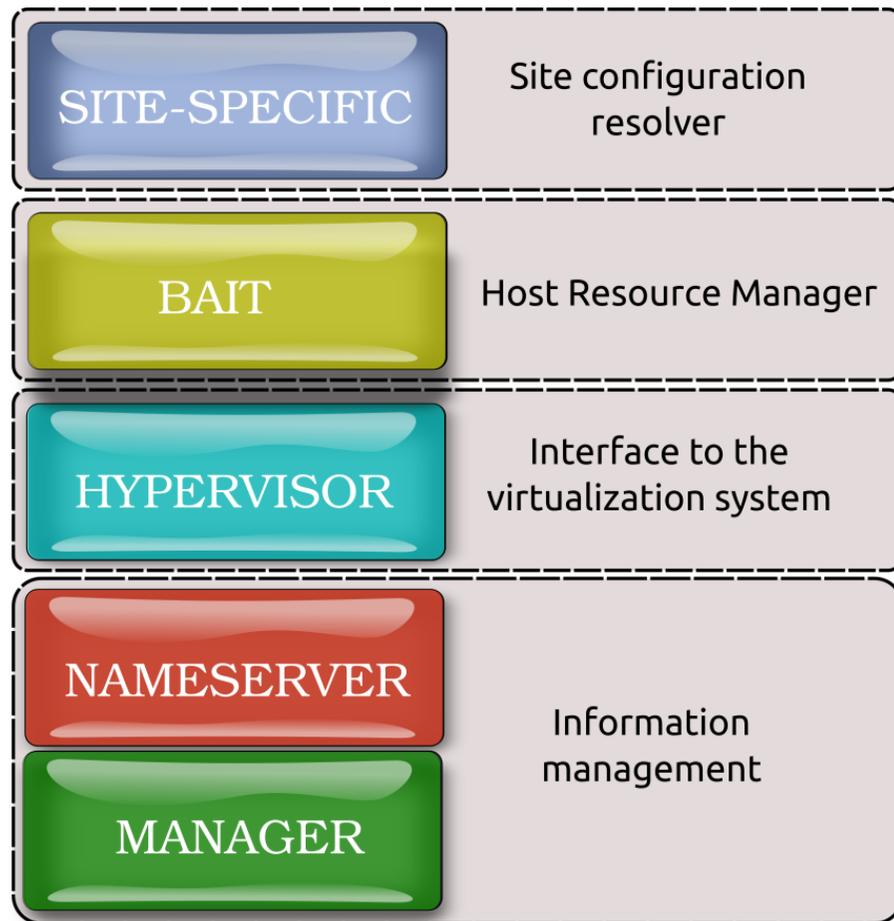
WNoDeS

- WNoDeS → **Worker Nodes on Demand Service**
- A software framework created by INFN to **integrate Grid and Cloud provisioning through virtualization**
 - **Key feature**: all resources (presented via Grid, Cloud, or else) are taken from a **common pool** to avoid static partitioning.
 - **Key feature**: resource matchmaking policies are handled by an LRMS.
- **Scalable and reliable** – it is **in production** at several Italian centers, including the INFN Tier-1 since November 2009
 - Currently managing about 2,000 on-demand Virtual Machines (VMs) there.
- **Transparent** for local and Grid Computing Center services
- Leveraging **proven open source software** technologies like Linux KVM, Torque/Maui (Platform LSF also supported), EMI gLite middleware
- **WNoDeS version 2 is part of EMI-2**, to be released this Monday (5/21)

WNoDeS, Architectural Overview



What's in WNoDeS 2

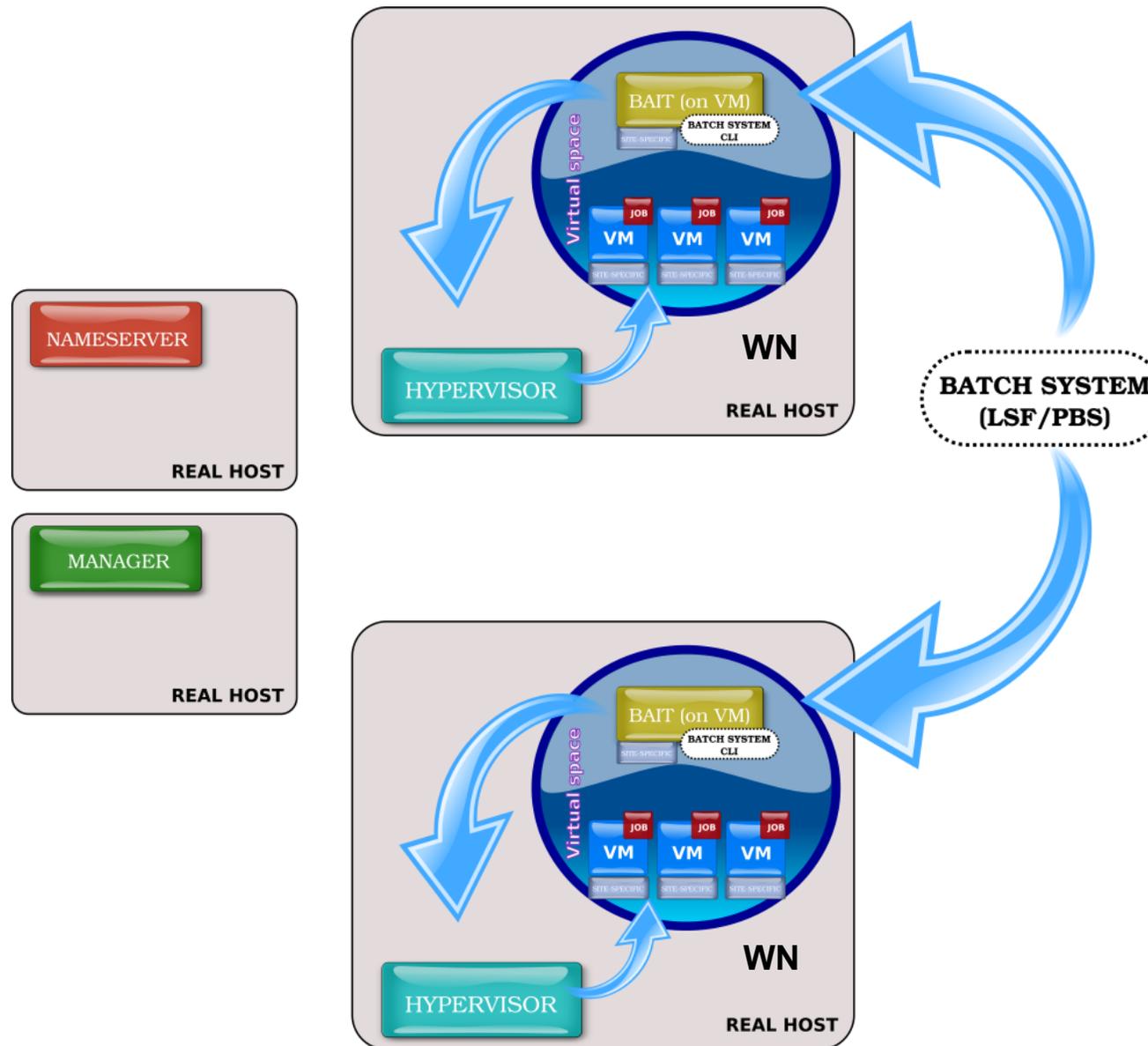


- WNoDeS 2 will be released this month **as part of the EMI-2 stack**
- **Batch job execution**
 - Local batch jobs can be run on both real and virtual execution hosts (worker nodes)
 - A virtual worker node is instantiated on-demand to provide computing resources as they are needed
- **Grid integration**
 - Dynamic selection of resources through standard gLite submission tools
 - Resource selection based on a Glue 1.x attribute
 - Authentication based on VOMS
- **Support for LSF and PBS**
- **New feature: mixed mode**

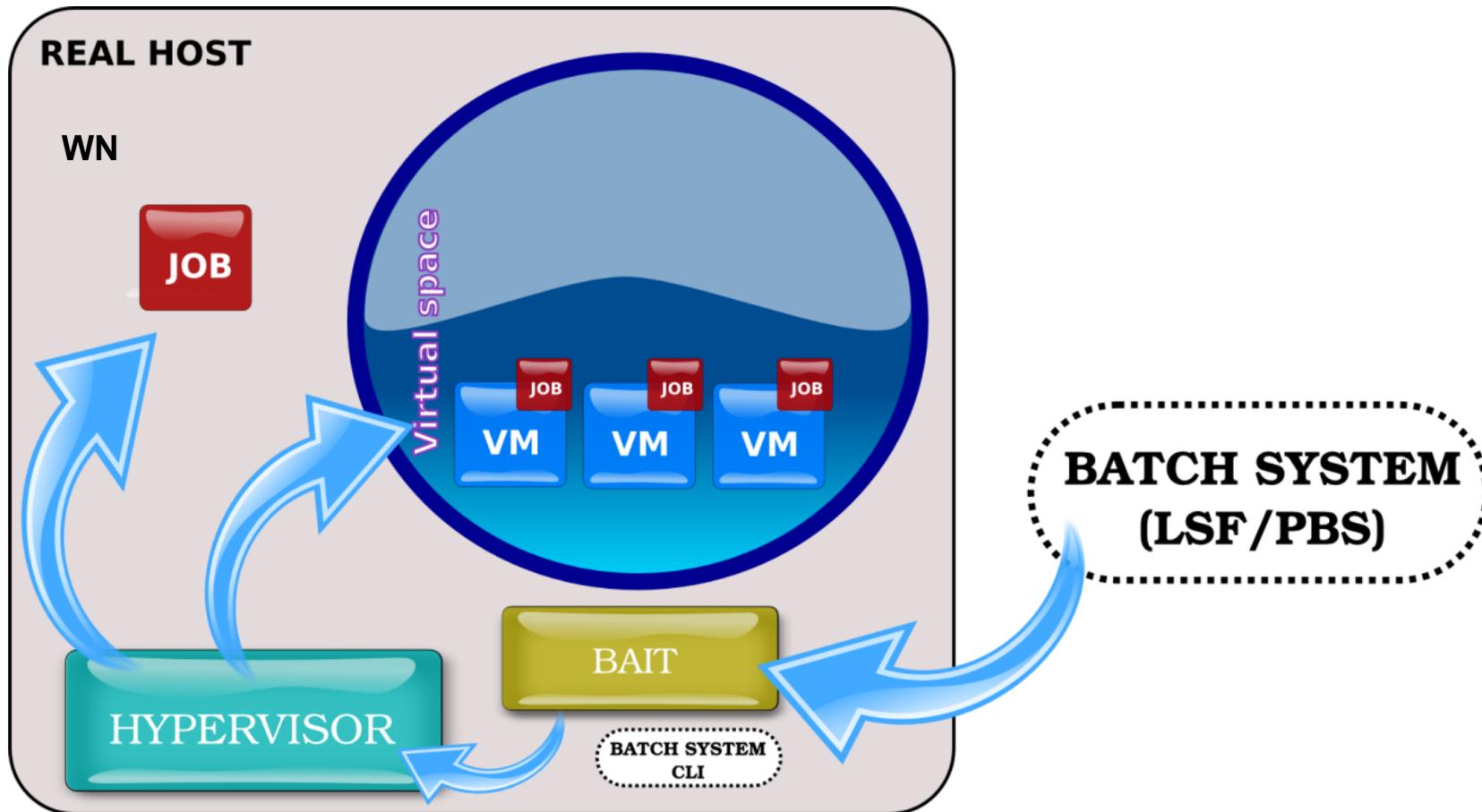
WNoDeS Mixed Mode

- **What**
 - A WNoDeS configuration option allowing the use of physical resources as both traditional batch nodes **and** as hypervisors for the instantiation of virtual machines – on the same hardware, at the same time.
 - VMs can be used to also run batch jobs, or to provide cloud services.
- **Why**
 - Some tasks are not suitable to being executed on virtual nodes – for example, jobs requiring GPGPU resources, or jobs with high I/O requirements → run them on physical nodes.
 - On the same physical nodes, one can also offer virtualized services for those users requiring them → no need to set aside nodes for virtualized services.
- **Where**
 - Mixed mode is included in the WNoDeS version released with EMI-2 and can be administratively turned on or off.

WNoDeS, Mixed Mode off



WNoDeS, Mixed Mode on



Mixed Mode Pros and Cons

- **Pros**
 - Progressively install WNoDeS in a farm without first having to decide which nodes will support virtualization and which not.
 - Add support e.g. for Cloud computing, interactive usage on custom VMs etc. in a traditional farm.
 - Direct jobs to VMs or to real hardware using LRMS policies and a simple pre-exec/prologue script (a template is supplied with the WNoDeS distribution). One can differentiate real vs. virtual requests/jobs e.g. based on queues, users, requirements, Grid VOs, etc.
- **Cons**
 - In a purely virtual farm set up, physical systems are only used as hypervisors, so they can be put e.g. in private address space. With mixed mode, they can also be used (like in a traditional farm) to run jobs and may need public access.
 - With mixed mode, a physical system is part of the LRMS cluster and may use up LRMS licenses proportionally to the number of its cores. If the same physical system is then also used to create VMs that become part of the LRMS cluster (e.g. to run batch or grid jobs), these VMs will also use up LRMS licenses and the total number of LRMS licenses used by a physical system may be $O(2 * \text{cores})$. This can be a problem with some sites using commercial LRMS.

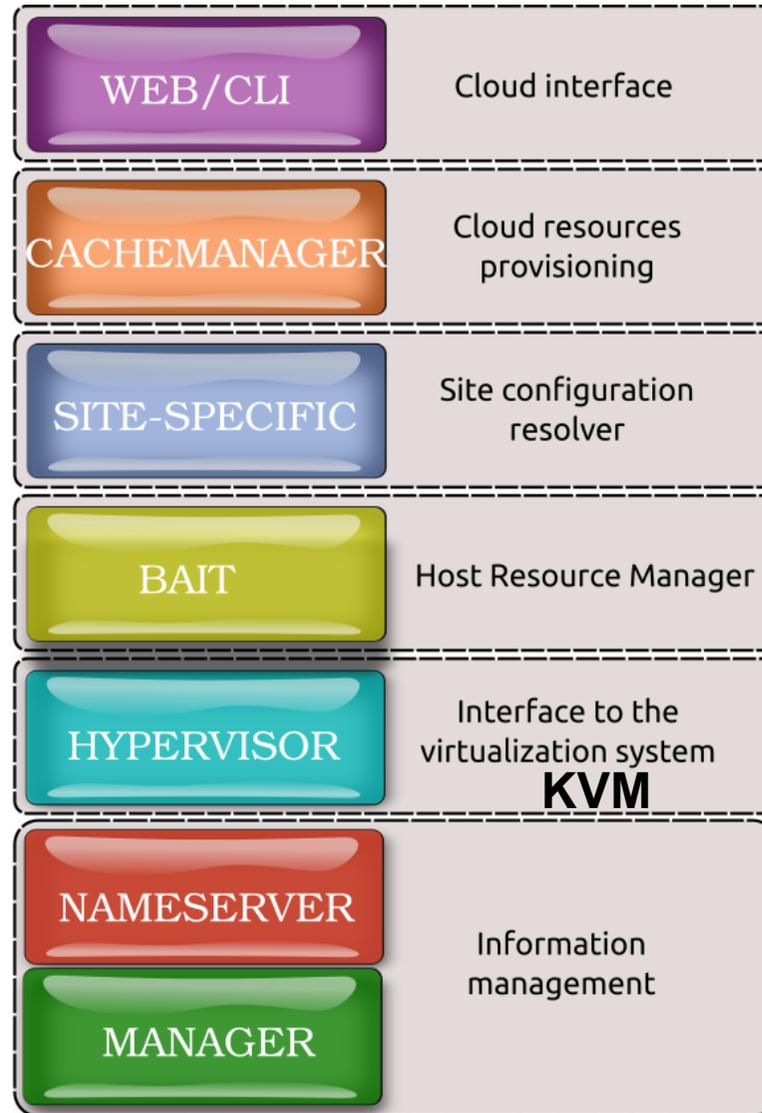
Some Upcoming Features (1)

- Several of the following features have been part of WNoDeS 1 (pre-EMI) for a long time. They will be released as EMI updates in the coming months.
- **Interactive usage**
 - We use this at the INFN Tier-1 to support self-instantiation of VMs by local users. These VMs can be used for e.g. analysis tasks, testing purposes, etc.
 - As with other WNoDeS services, resources can be taken from the general purpose Tier-1 farm (no service partitioning, unless one specifically configures it.)
- Cloud computing: **OCCI compliance**
 - Will be compliant to OCCI 1.1 (validated e.g. through doyou speakocci) and accessible via a CLI.
- Cloud computing: **Web interface**

Some Upcoming Features (2)

- **Dynamic virtual networks**
 - Won't require using 802.1q to partition networks and will allow dynamic instantiation of private VLANs (either local or across multiple sites) and address assignment for VM isolation – a much needed feature in cloud environments.
 - Also allowing possible provisioning of services like *Cloud bursting* (to other resource centers), or *Cluster as a Service*.
- **Integration of multiple authentication methods**
 - The current WNoDeS cloud Web application and OCCl interface use X.509 + VOMS; this will be extended to support federated access.
- **Storage volume support**
 - Will allow dynamic instantiation and connection to VMs of persistent storage volumes.

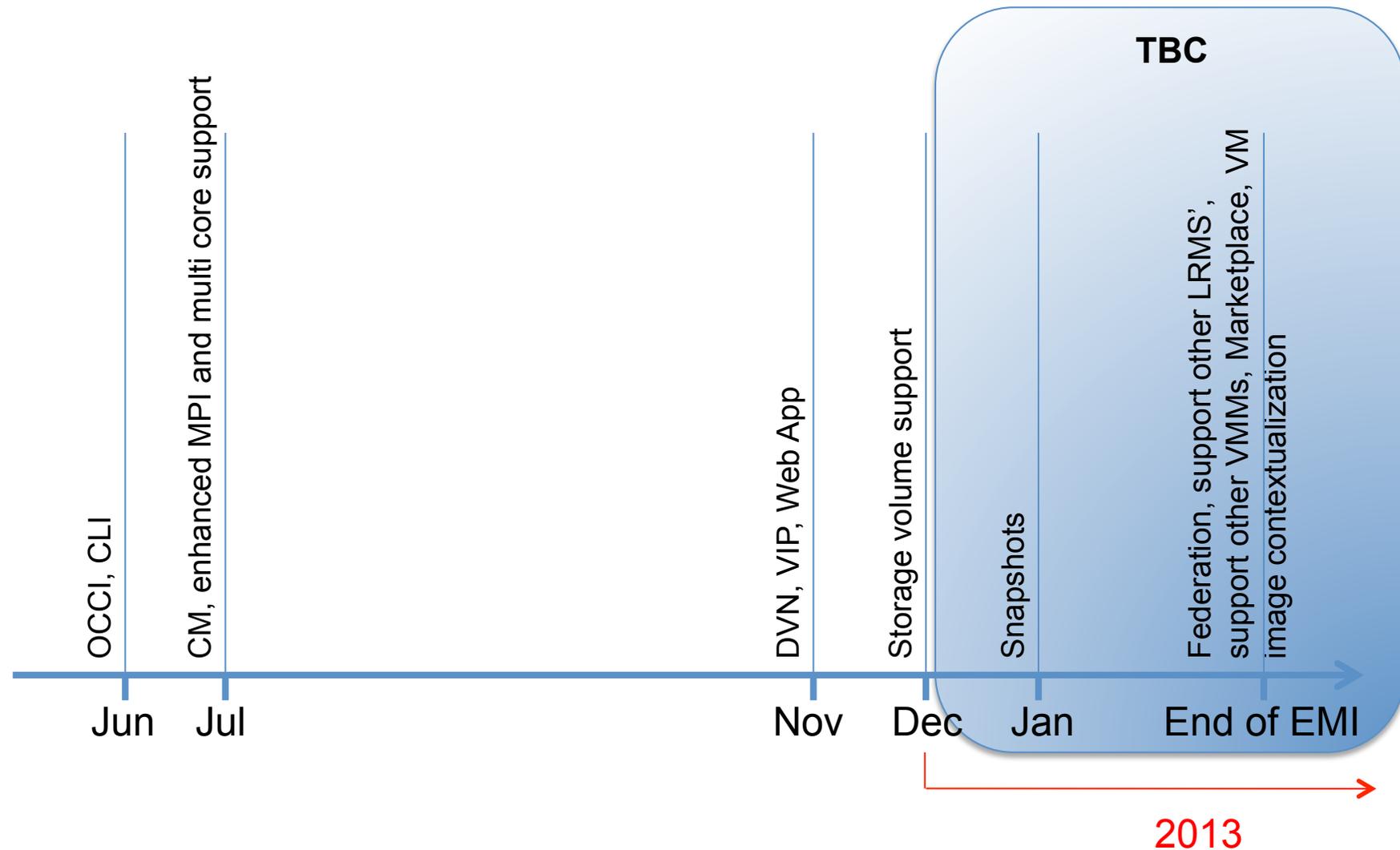
WNoDeS, Cloud Support



EMI2 Updates

EMI2

Current WNoDeS Timeline



Other R&D Areas at CNAF

- Infrastructure
- Storage
- Farming
- Network
- Testing
- National / International Projects

Infrastructure

- Power management / green computing
 - Parameter optimization for the data center
 - Testing of software packages integrating management of infrastructure and computing resources; development of extensions to handle farming control (stand-by / shut-down, power savings modes)

Storage Area

- Tiered storage: storage services differentiated wrt quality, analysis of redundant technologies
- Evolution and performance of shared filesystems (GPFS, NFS 4.1, GFS, GlusterFS)

Farming Area

- Batch system evolution (LSF, PBSPro, Torque, SLURM)
- CPU performance: testing of new platforms, analysis of some implementations (e.g. Turbo Boost / Turbo Core, Hyper-Threading)
- Optimization / extension of batch system configurations
- MPI support

Network Area

- Testing of solutions alternative to what we currently have wrt to data transmission (FCoE, iSCSI)
- Dynamic circuits on WANs (Software Defined Networks)

International / National Projects

- Participation to European Distributed Infrastructures (European Grid Infrastructure [EGI], European Middleware Initiative [EMI], Standards and Interoperability [SIENA])
- Italian Grid Initiative
- SuperB: definition of the computing framework
- Several IT-financed projects ongoing
- Computing on Knights Architecture (COKA)
 - Evaluation and porting of apps on the Intel MIC platform (in collaboration with Intel and several HPC centers)

Thanks

- More info:
 - <http://web.infn.it/wnodes>
 - wnodes-support@lists.infn.it
 - The WNoDeS software will be available and supported through the usual EMI channels after EMI-2 is released.