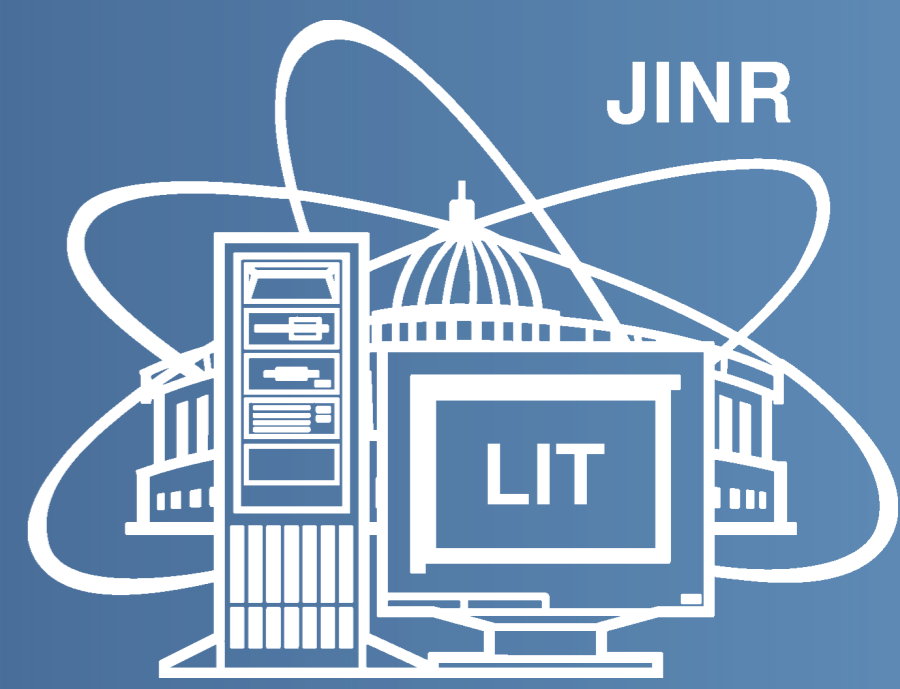


VM-based infrastructure for simulating different cluster and storage solutions used on ATLAS Tier-3 sites



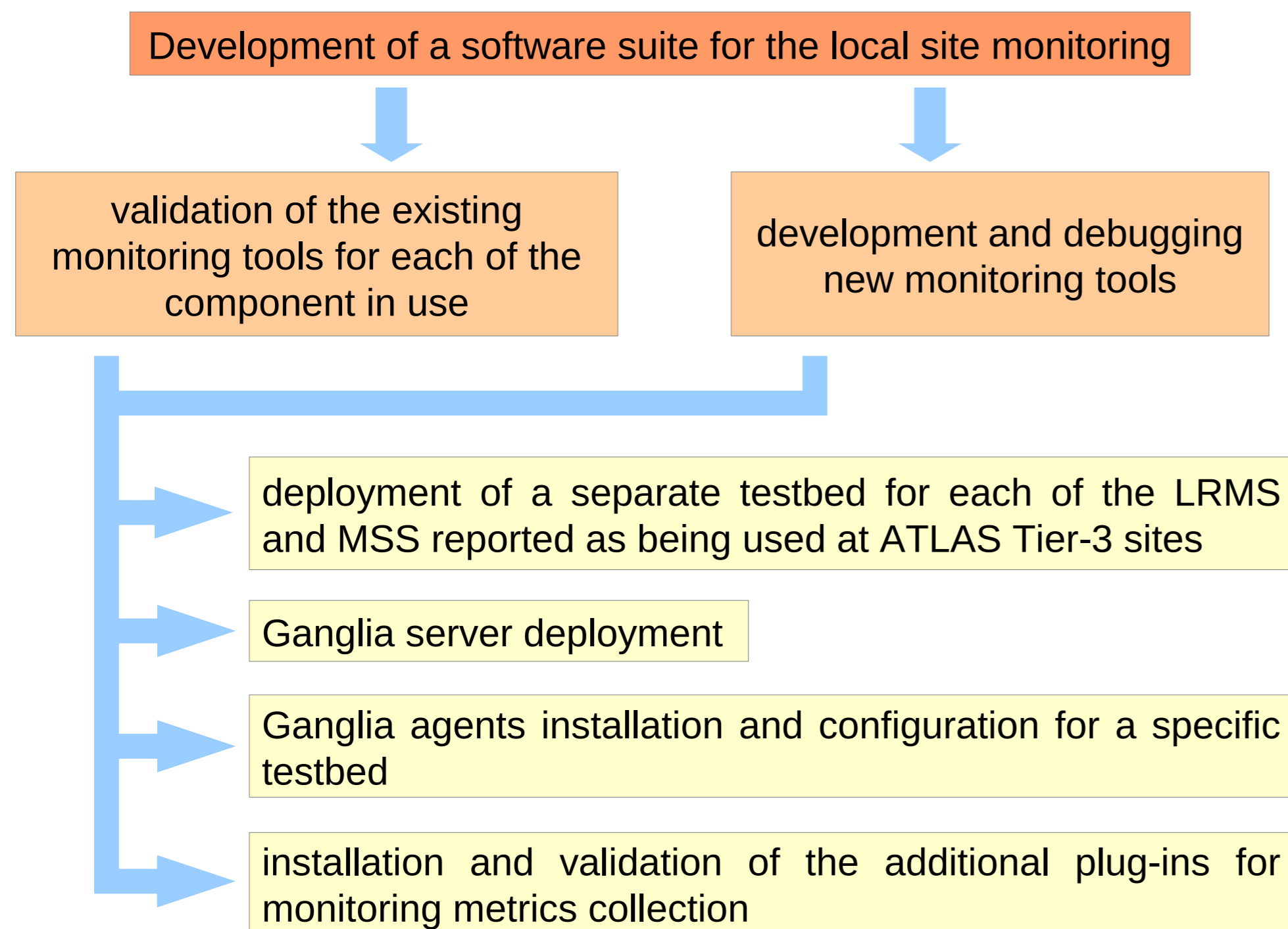
LRMS & MSS types on Tier-3 sites

According to ATLAS Tier-3 survey [1] a several types of the local resources management systems (LRMS) and the mass storage systems (MSS) are used on the ATLAS Tier-3 sites (see Table 1).

Table 1. A list of the LRMS and MSS used at ATLAS Tier-3 sites

LRMS	PROOF, PBS, Condor, Oracle Grid Engine, LSF
MSS	XRootD, dCache, DPM, NFS, GPFS, Lustre

Tasks



Virtualization

- 24/7 availability of the testbeds components with different LRMS and MSS running in parallel on dedicated physical servers would causes a sufficient hardware capacities
- monitoring tools deployment and development as well as testbeds operation may require redeployment of a certain testbed or its parts
- testbeds performance is not a critical issue for such tasks

Virtualization:

- more effective utilization of the hardware resources,
- ability to perform quickly and easily such operations as VMs creation from existing images/templates, VMs backup before significant changes and VMs restoration from backup if needed.

- all components of each testbed can be run on linux (inside VM) as well as a physical servers;
- most components do not require own kernel extensions and thus the OS-level virtualization can be used which is more lightweight and faster than full hardware emulation or paravirtualization approaches;
- but there are still some components which require own kernel extensions (e.g. Lustre, GPFS).



The following hypervisors were chosen:

- **OpenVZ** (as a solution for virtualization on OS-level),
 - **Xen** (as a hypervisor implementing full hardware virtualization),
- due to the reasons listed below:
- stable and actively developing software with sufficient tool set for VMs management and monitoring,
 - strong and helpful community,
 - good documentation,
 - free software (GNU GPL license),
 - own successful long-term experience of their usage.

Testbed structure

A scheme of LRMS and MSS testbeds distribution over the physical servers is shown on Figure 1 and a list of the services running on each component of the deployed testbeds is given in Table 2.

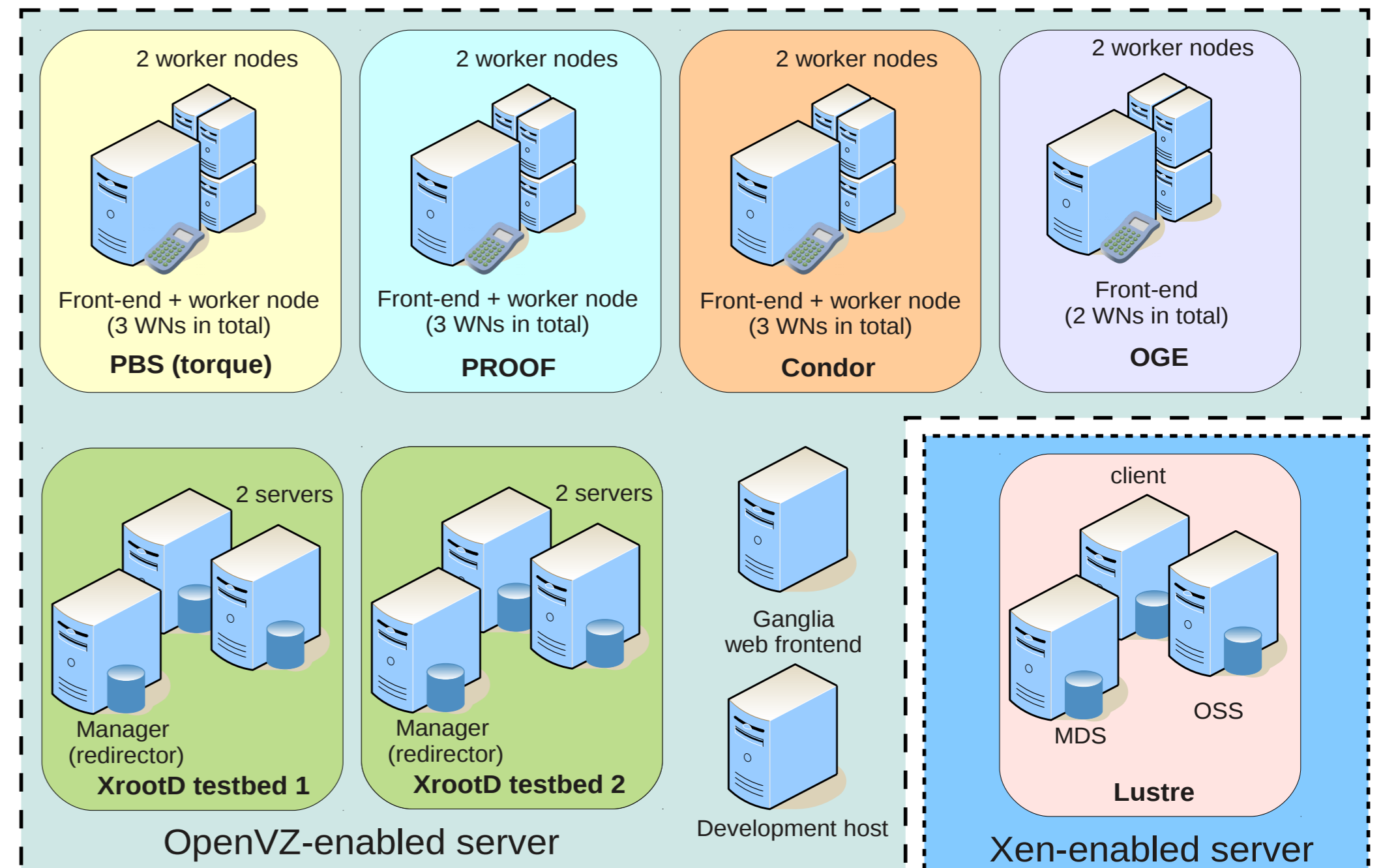


Figure 1. A scheme of the LRMS and MSS testbeds distribution over the servers

Table 2. A list of running services of the LRMS and MSS testbeds

Testbed	Services
PBS	torque headnode (HN) + worker node (WN) + ganglia (gmond, gmetad, webfronted) + jobmonarch, 2 torque WNs + gmond
PROOF	HN + gmond, 2 WNs + gmond
Condor	HN + WN + gmond, WN + gmond, client + gmond
OGE	HN + Ganglia (gmond, gmetad, webfrontend), 2 WNs + gmond
XRootD 1	manager + gmond, server + gmond, server + ganglia (gmond, gmetad, webfrontend)
XRootD 2	manager + gmond, server + gmond, server + ganglia (gmond, gmetad, webfrontend)
Lustre	MDS + gmond + gmetad + gweb + t3mon-site-lustre, OSS + gmond, client + gmond

Load test suite

To develop reliable and valid monitoring tools some activity needs to be simulated on the deployed testbeds.

- Job events
 - random submissions with configurable frequency,
 - adjustable memory usage,
 - CPU load.
- File events
 - uploading file to storage (random size, random time),
 - remote file existence check,
 - deletion of the file after some time.

For that purpose the load test suite has been developed.

- Currently supported MSS and LRMS: XrootD, Condor, Torque, OGE.
- Event parameters:
 - start time, file parameters, job memory usage, etc. have a uniform distribution by default, maximum values could be adjusted to the cluster's configuration.
- Instrumentation:
 - python library and bash, command line interface,
 - event series generation is started by cron.

References

1. D. Benjamin "ATLAS Tier-3 survey" / Report at ATLAS Software and computing workshop (29.11.2010 – 03.12.2010), CERN (access is granted for ATLAS collaborators only).