Tier 3 Monitoring Software Suit (T3MON)

S.D. Belov[1], I.S. Kadochnikov[1], N.A. Kutovskiy[1,2], D.A. Oleynik[1], A.S. Petrosyan[1], L. Valova[1,3]

[1]Laboratory of Information Technologies, JINR, Dubna

[2] National Scientific and Educational Centre of Particle and High Energy Physics of the Belarusian State University, 220040, Minsk, Belarus

[3]Technical University of Kosice, 040 11, Kosice, Slovakia

Introduction

The ATLAS Distributed Computing activities concentrated so far in the "central" part of the computing system of the experiment, namely the first 3 tiers (CERN Tier0, the 10 Tier1s centers and the 60+ Tier2s). This is a coherent system to perform data processing and management on a global scale and host (re)processing, simulation activities down to group and user analysis.

Many ATLAS Institutes and National Communities built (or have plans to build) Tier-3 facilities. Tier-3 is a non-pledged computing resource for local scientific communities (Institute/University) with a limited local and centralized support. Tier-3 centers consist of resources mostly dedicated for the data analysis by the geographically close or local scientific groups. Tier-3 sites comprise a range of architectures and many do not possess Grid middleware, which would render application of Tier-2 monitoring systems useless [1].

In March 2011 a proposal was approved, which describes a strategy of development monitoring software for non-pledged resources: Tier-3 Monitoring Software Suite (T3MON) [2]. T3MON software package should meet the requirements of the ATLAS collaboration for global monitoring of ATLAS activities at Tier3 sites, and the needs of Tier3 site administrators. The solutions implemented in frames of this project are expected to be generic, so other Virtual Organizations (VO), within or outside of LHC experiments, can use them.

Software suite:

- will allow a site administrator to monitor local Tier3 fabric(s);

- will provide a global monitoring view to the services provided by the Tier3 center, namely:

  - data transfers to the site and between sites;

  - data processing and analysis.

Main components of the system are:

- a software suite for local site monitoring – "T3MON-SITE";

- information system which will aggregate and visualize data from distributed Tier3 sites at a global VO level – "T3MON-GLOBAL".

The main requirements for "T3MON-SITE" package are simple installation and support, intuitive user interface. The package should provide a low level monitoring of all site resources, status and performance of the hardware components, activities of the VO at the site. The toolkit should also include monitoring of data files located at the site. The toolkit should foresee a possibility for propagation of the aggregated monitoring metrics of the VO activities at the site to the VO central Tier3 monitoring system ("T3MON-GLOBAL").

Central Tier3 monitoring should be based on data collected from the local monitoring systems at Tier3 sites. This data contain aggregated monitoring metrics of VO job processing and data transfer at a given Tier3 site. The service must be scalable and has a minimal impact on the local resources. The set of the monitoring metrics as well as its granularity have to be defined by ATLAS.

Implementation of "T3MON-SITE"

In light of the results of Tier3 survey [3] and in accordance with the requirements, we developed a package based on the Ganglia [4] monitoring system. Ganglia is an open-source package used for real-time monitoring of large UNIX clusters. Each node in a Ganglia system runs a daemon that reports on the state of its host in the form of performance metrics including memory, CPU, load, disk and network statistics. Collectors gather data produced by the daemons and store it in round-robin databases. The information is typically presented in the form of plots via a webserver, but can be also obtained in XML format and consumed by various clients.

The main development effort was concentrated on enabling plug-ins for PROOF [5] and XRootD [6] monitoring through Ganglia.

JINR Tier3 testbed [8]

The development of software suite for local site monitoring assumes the following activities:

- validation of existing monitoring tools for each of the using component,

- development and debugging the new monitoring tools.

The activities listed above imply the following:

- deployment of separate testbed for each of the LRMS and MSS reported as being used at ATLAS Tier3 sites,

- Ganglia and Nagios servers deployment,

- Ganglia or/and Nagios agents installation and configuration for a specific testbed,

- Installation and validation of the additional ganglia plug-ins for monitoring metrics collection as well as non-related to ganglia monitoring tools.

For estimating this aims special testbed based on VM technology was deployed in JINR.

Implementation of "T3MON-GLOBAL"

The central Tier3 monitoring system is based on monitoring data published by Tier3 sites and should provide a global picture of how ATLAS uses Tier3 resources. The necessary condition for the development of the central Tier3 monitoring system is consistent registration of the Tier3 sites in the ATLAS Grid Information System (AGIS). Another important factor is encouraging the Tier3 user community to use data transfer and job submission systems which are instrumented for reporting the monitoring data, for example Ganga, Athena, dq2.

The system consists of several components:

- Publishing agents, which will run at Tier3 sites, will interact with the local monitoring systems, aggregate and publish monitoring metrics to the message bus. As a transport layer, we plan to use the Apache ActiveMQ messaging system. Apache ActiveMQ is an open source messaging system which was recently evaluated as a standard messaging solution for the WLCG infrastructure.

- Data collector from the ActiveMQ message broker. Collected data will be recorded in the central data repository (ORACLE backend).

- Data presentation layer includes an interactive user interface and an API for data export. Most probably, the additional Tier3 views will be enabled in the existing Dashboard ATLAS monitoring systems, namely ATLAS DDM Dashboard and ATLAS Global job monitoring.

Current status of project

The software part consists of XRootD and Proof plug-ins for Ganglia developed.

Implementation of "T3MON-SITE" enables the monitoring of the local infrastructure for site administration:

- detailed monitoring of the local fabric (overall cluster or clusters monitoring, monitoring each individual node in the cluster, network utilization);

- monitoring of the batch system (distribution of tasks on nodes);

- monitoring of job processing (PROOF);

- monitoring of the mass storage system (total and available space, number of connections, I/O performance).

It also includes:

- Test cluster installation and configuration was performed, PBS, Torque, LSF, Condor, XRootD and PROOF clusters were installed. Testbed is based on virtualization technology, at the moment there are 18 virtual machines, all run on one physical server (AMD Athlon 64x2 Dual Core 3800+, 4Gb RAM, 320Gb HDD). Test load suite provides:

  - 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 configured period of time.

- Ganglia packages were installed at each cluster; all clusters are being shown via one main web interface [11].

- Document installation and configurations of Ganglia for various cluster configurations [12].

- Development of PROOF and XRootD (EOS) add-ons was performed.

  - XRootD/Ganglia monitoring software was integrated with T3MON suite, and installed on volunteer sites, such as tier 3 site at JINR, BNL, IHEP, and in Ukraine;

  - EOS data popularity plug-in is now being tested on CERN EOS cluster;

– PROOF add-on is in testing stage.

"T3MON-GLOBAL" part is in development stage:

- metrics list was offered, discussed and agreed by the collaboration;

- publishing agents are being developed.

The Tier-3 monitoring task force was presented on ATLAS Software & Computing Workshop in April [7], testbed at JINR and XRootD monitoring for local site solution on ATLAS Technical Interchange Meeting [8]-[9], a status report was presented on ATLAS Software& Computing Workshop in July 2011 [10].

References

[1] Brock R. et al. U.S. ATLAS Tier 3 Task Force. Preprint U.S. ATLAS, 2009.

[2] Andreeva J. et al. Tier-3 Monitoring Software Suite (T3MON) proposal. ATLAS note, 2011.

[3] Benjamin D. Tier 3 Survey results. Report at ATLAS Software & Computing Workshop (29 November 2010 - 03 December 2010), CERN, Geneva.

[4] Ganglia monitoring system, http://ganglia.sourceforge.net/

[5] PROOF, http://root.cern.ch/drupal/content/proof

[6] XRootD, http://xrootd.org/

[7] Petrosyan A. et al. Tier 3 Monitoring Task Force. Report at ATLAS Software& Computing Workshop (5 April 2011 8 April 2011), CERN, Geneva.

[8] Kutovski M. JINR infrastructure for Tier 3 simulation. Report at ATLAS Computing Technical Interchange Meeting (31 May 2011 2 June 2011), JINR, Dubna.

[9] Kadochnikov I. et al. XRootD monitoring for local site. Tutori al. Report at ATLAS Computing Technical Interchange Meeting (31 May 2011 2 June 2011), JINR, Dubna.

[10] Petrosyan A. et al. Tier 3 monitoring. Report at ATLAS Software& Computing Workshop in July 2011.

[11] T3MON project home, https://svnweb.cern.ch /trac/ t3mon/wiki/T3MONHome