How to integrate portals with the EGI monitoring system

Dusan Vudragovic

Scientific Computing Laboratory

Institute of Physics Belgrade, Serbia

- Introduction
- Overview and initial proposal for integration of Scientific Gateways (SG) into
  - GOC DB
  - GStat
  - SAM Framework
  - Real Time Monitoring
  - GridView
  - Google Earth
  - GridMap
  - Operations Portal

- Scientific Gateways have become an essential tool for research
- Their operation and performance has to be monitored in order to ensure quality of service for end-users
- Such monitoring has to ensure an integrated overview of the global status of scientific gateways, but also detailed status of the individual scientific gateway layers and components
- In addition to this, the monitoring has to:
  - enable sending of alerts to administrators when a particular problem is identified
  - enable scheduling of downtimes during SG maintenance
  - produce SG performance statistical reports

- Currently, several monitoring tools are used by EGI to detect and diagnose problems with sites
  - GOC DB
  - GStat (GIIS monitoring)
  - SAM framework
  - Real Time Monitoring
  - GridView
  - Google Earth
  - GridMap
- On top of these, Operations Dashboard provides links and utilizes combined views to simplify monitoring tasks

- Central static information repository
- Stores information about NGIs, sites, nodes, services, users, etc.
- Used to declare maintenance for (un)scheduled events
- Consists of three parts:
  - database, where all information is stored
  - web portal, which interfaces with the database
  - programmatic interface
- Exports initial configuration for the information system

- Definition of a new object in GOC DB – Scientific Gateway
- Definition of the attributes of this object
  - name of the portal
  - portal URL
  - type of the portal
  - version of the portal
  - contact persons (sysadmin, user support, security)
  - available applications, etc.
- Definition of SG LDAP URL (in this way, SG can dynamically publish information to the Grid information system)

GStat

- Visualizes Grid infrastructures from an operational perspective, based on information found in the Grid information system

- Checks the health of Grid information system:
    - detects faults in the information system
    - verifies the validity of information

- This is done by directly querying site information systems and top-level information system

- It periodically takes a snapshot of the information system and maintains a cache of the main entities found in the infrastructure

- GStat can provide statistical information on all properties available in the information system
  - Number of jobs (total/running/waiting)
  - Number of jobs per application
  - Number of available job slots
  - Number of users
  - Available applications

- Relies on existing technologies
  - Nagios is used for scheduling and execution of the probes
  - MSG messaging system (ActiveMQ) integrates other operational tools with Nagios instances
- SAM framework provides:
  - Status and history of services and sites
  - Visualization of services and sites' availabilities
  - Web services for data exports
- Nagios has a pluggable architecture that allows easy integration of SG probes
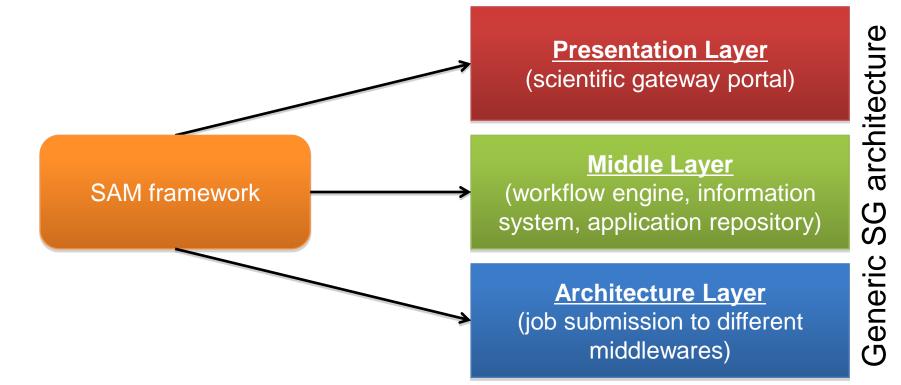
- SAM uses three central databases:
  - Aggregated Topology Provider (ATP)
  - Metric Description Database (MDDB)
  - Metric Results Store (MRS)
- Nagios Config Generator (NCG) enables automatic generation of Nagios configuration based on multiple information sources
- Nagios Probes
  - Simple Probes (check of a service in a single run)
  - Multitest Probes (single run performs multiple tests; mix of active and passive checks; file put > file get > file delete)
  - Long-running Probes (submit > monitor > report state)

- ## SAM framework has to ensure monitoring of all SG layers and components

Presentation Layer
(scientific gateway portal)

Middle Layer
(workflow engine, information system, application repository)

Architecture Layer
(job submission to different middlewares)

SAM framework

Generic SG architecture

- Probes for the Presentation Layer
  - availability of portal and its components
  - check of the authentication mechanism
  - check of the input data management
  - check of the workflow and data-flow tool
  - application submission
- Probes for the Middle Layer
  - application repository checks
  - check of the workflow storage and interpreter
  - check of the local file storage
- Probes for the Architecture Layer
  - check of the submission to different DCIs (gLite, ARC, Unicore, Globus, LFS, PBS, BOINC, web service, local resource, Google App Engine, etc.)

- Real time monitor overlays Grid activity onto a 3D globe
- Each Grid site is represented by a circle at the location of the resources (pulsing circle of magenta and green)
- Workload Management Systems (WMS) are represented as triangles
- Special symbol can be assigned for representation of SGs, while the number of jobs might be retrieved from the information system or through the RTM-dedicated service at SGs

- GridView visualization tool provides a high-level view of various functional and performance aspects using GridFTP, WMS, FTS logs and SAM MSG

- Google Earth geographical location of sites

- GridMap gives graphical representation of site CPU power and its availability

- Entry point for all information and services related to EGI's operations, where the community can manage, monitor, share and discuss information

- Architecture
  - database – to store information related to the users and VOs
  - web module – graphical user interface
  - data aggregation and unification service

- In addition to COD and VO dashboards, a new SG dashboard could be introduced
  - dashboard with the overview of all detected problems related to SGs
  - enabling operations staff to track problems using different results from various monitoring tools
- VO info feature could be also provided for SGs
  - information on how to support (offer resources) a particular SG
- Broadcast feature
  - contact several categories of stakeholders interested in notifications about identified problems, issues, downtimes
  - announcement of a new SG version, or a new application