Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

Forschungszentrum Karlsruhe GmbH, Institute for Scientific Computing,
Postfach 36 40, 76021 Karlsruhe

Andreas Heiss, Bruno Hoeft, Axel Jaeger, Holger Marten, Bernhard Verstege

Monitoring a WLCG Tier-1 Computing Facility
aiming at a reliable 24/7 service

The mission

Provide resources and services for LHC experiments and others with a high reliability and availability.

- Accept raw and reconstructed data for storage on disk and tape.
- Provide data to other Tier-1 and Tier-2 sites with high rate.
- Accept MC generated data from associated Tier-2 sites.
- Ensure high-capacity wide area network bandwidth.
- Operation of a data-intensive analysis facility.
- Provide (gLite) Grid services (CE, SE, LFC, FTS, BDII, RB, PX,...)
  Critical services other sites (Tier-2) depend on:
  - *File Transfer Service (FTS)*
  - *Catalogues (LFC)*
  - *Information System*
  - *SRM / storage*

The challenge

- ≈ 1000 worker nodes (2500 CPU cores)
- ≈ 250 servers (file server, dCache pool nodes, login nodes, gLite servers etc.)
- almost 2 PB of disk for dCache and GPFS
- 1.5 PB of tape capacity

Numbers will more than double in 2008!

- complex SAN environment
- complex network setup (5 routers, ≈ 70 switches, VLANs, firewalls, ...)

- need to keep machines and services running 24/7
- need to know the dependencies between different services
- need to react fast and properly in the case of failures

The approach

- Thorough monitoring of machines and services:
  - use existing monitoring tools proven to scale.
  - do not ”re-invent the wheel”, but adapt tools to own needs.
  - provide single entry point to all monitoring information to get quick a complete picture of the overall situation.
- Automatise recovery procedures where possible *(see →Nagios)*
- Have admins/operators on-call outside working hours.
- Have experts on-call for critical services.
  → almost impossible to ensure anytime availability of experts for all services!
  → provide tools and recipees for non-expert personnel to further investigate and fix problems.
  → build up expertise of non-expert people to solve typical problems.

Service Availability Monitoring
Results of **S**ite **F**unctional **T**ests

Datasource:
http://lcg-sam.cern.ch:8080/sqldb/....

Network monitoring
- Cacti
- Router log file analysis
  (to be implemented)

Batch system information
- Job status
- number of jobs per VO
- cpu time / wall time ratio of jobs
- ...

FTS monitoring
Statistics on successful and failed transfers to each site.

Monitoring scripts are kindly provided by Matt Hodges (RAL) and Ron Trompert (SARA).

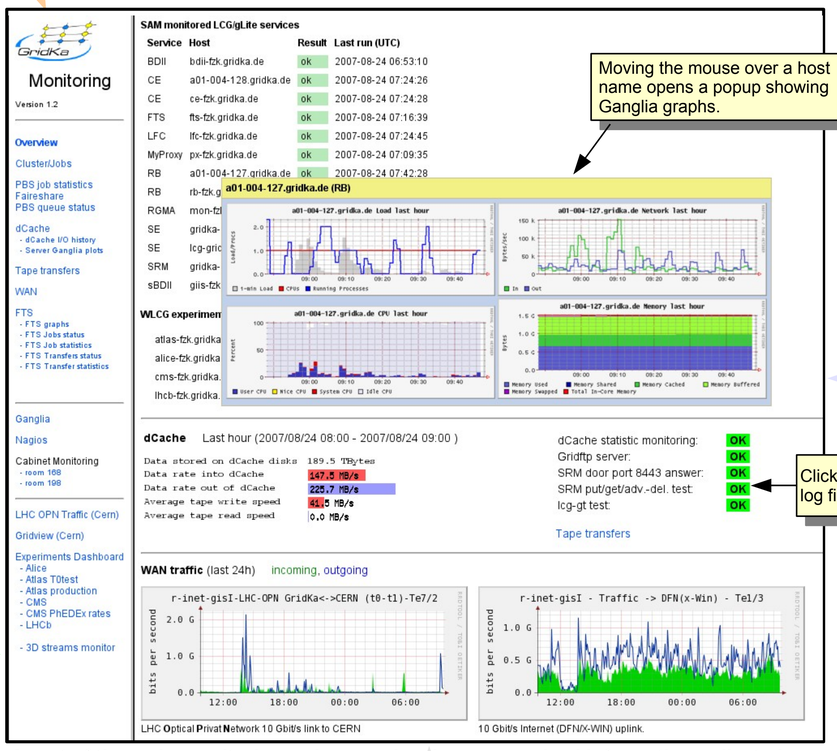
Results are fed into Nagios and displayed on web page.

Results are fed into Nagios.
Cacti graphs are displayed on web page.

Results are fed into Ganglia and displayed on the monitoring web page.

Statistics on transfer jobs are fed into Ganglia and displayed on the monitoring web page in differenr forms.

Nagios is used as a machine and service monitoring tool and is the central alarming system. At GridKa, more than 60 different checks, e.g. ping times, disk usage, log-file sizes, response times, SFT results, dCache functional tests, temperatures etc., are performed by Nagios. It is planned to feed information from the experiments' own monitoring into Nagios. As of today, ≈1650 hosts and ≈14000 services are monitored. Alarm notifications can be issued via email and mobile phones and recovery procedures (e.g. reboot a workernode) can be triggered automatically.

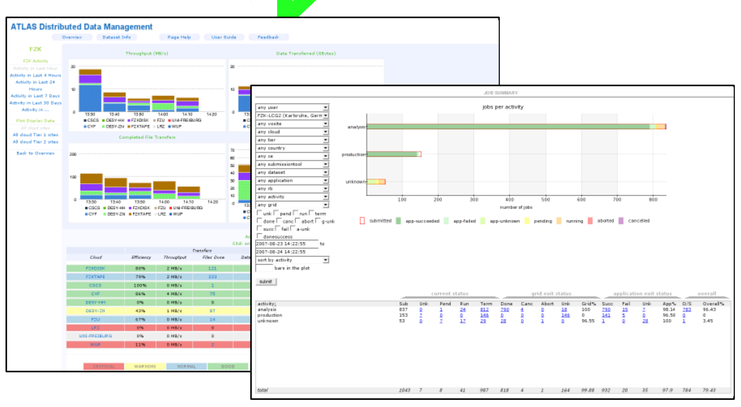


A part of the main monitoring web page showing summarized information collected from different sources, e.g. Ganglia, SAM, dCache tests, Cacti (routers). Subpages and linked web pages provide more detailed information.

Ganglia is used to monitor and visualize machine load, memory utilization, network load and other metrics.

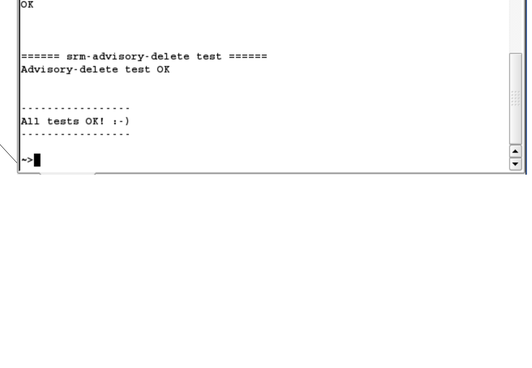
Monitoring data from the experiments dashboards will be fed into Nagios. An overall “health” indicator will be displayed for each experiment. (This is not yet available and has to be implemented.)

Test results are pushed to public web server and fed into Nagios.



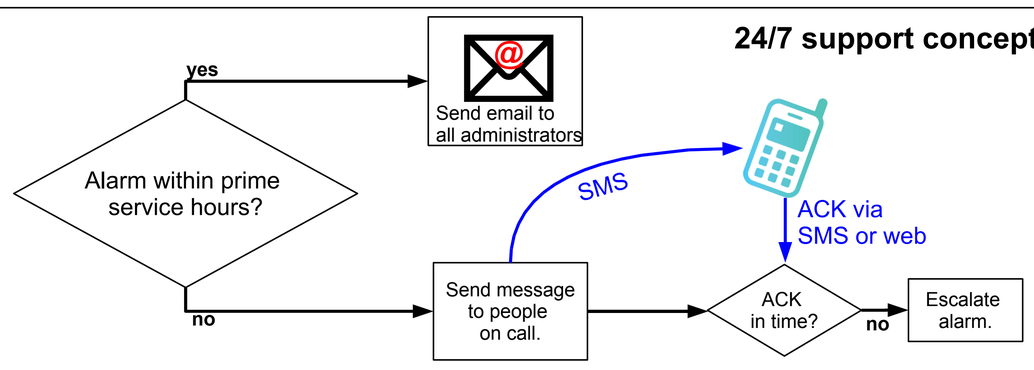
Monitoring server

- Runs automated tests on
  - dCache components: gridftp, SRM, ...
  - FTS
- Runs SFTs *(to be implemented)*
- Provides easy to use scripts for 'operators' to start tests manually and get detailed log files.
- Provides a gLite UI.
- Generates and updates web pages on public web server.



Nagios event on critical service

24/7 support concept (preliminary)



- **Planned on-call groups:** (≥ 2 people each)
- **network** ✓
- **infrastructure (power, cooling)** ✓
- **storage**
- **grid services**

Andreas Heiss
Phone: +49 7247 82-5639
Email: andreas.heiss@kit.edu
Web: www.gridka.de