CONNECTING WLCG TIER-2 CENTRES TO GRIDKA

A. Heiss, S. Halstenberg, B. Hoeft, D. Ressmann, J. van Wezel,
Forschungszentrum Karlsruhe, Germany

Abstract

GridKa, the German Tier-1 centre in the Worldwide LHC Computing Grid (WLCG), supports all four LHC experiments, ALICE, ATLAS, CMS and LHCb as well as currently four non-LHC high energy physics experiments. Several German and European Tier-2 sites will be connected to GridKa as their Tier-1. We present technical and organizational aspects pertaining the connection and support of the associated Tier-2s sites. These include networking and security issues as well as the storage system layout. The storage system layout at GridKa has to ensure that the data stream from the Tier-0 to GridKa can be written at all times and is not interfered by other incoming and outgoing data streams.
First results of file transfer tests between GridKa and some of the Tier-2 sites, using the currently available shared network links, are shown.

GRIDKA AND ITS ASSOCIATED TIER-2 CENTRES

The Worldwide LHC Computing Grid (WLCG) is organized in a multi-tier structure with CERN, the Tier-0, as the main data source. The Tier-1 centres have multiple responsibilities, such as storing and processing raw data, providing reconstructed "ESD/RECO/DST" and "AOD" data to other Tier-1 and Tier-2 centres as well as storing Monte Carlo (MC) simulated data produced at Tier-2 sites. The Tier-2 centres are mainly used for user analysis and MC generation, however this depends on the computing models of the experiments. In comparison to Tier-1 centres Tier-2 sites do not need large mass storage systems, in particular tape storage is not a requirement at Tier-2 sites. GridKa, located at Forschungszentrum Karlsruhe, Germany, is a Tier-1 centre for the four LHC experiments Alice, Atlas, CMS and LHCb within the WLCG project. The GridKa resources are also used by several non-LHC experiments, CDF, D0, Barbar and Compass, which process and store part of their data there. The share of the resources dedicated to the LHC experiments is currently $\approx 50\%$ but will increase to $> 90\%$ in 2008 and later.

Multiple Tier-2 centres, located in Germany and other European countries, which support one or more of the four LHC experiments, will use GridKa as their associated Tier-1 site. In contrast to the WLCG tree structure with several Tier-1 centres connected to the Tier-0 and Tier-2 centres connected to each Tier-1, the EGEE[2] resource centres are grouped in "regions", e.g. Germany/Switzerland (DECH) where GridKa belongs to. As opposed to other WLCG Tier-1 sites, GridKa will have associated Tier-2 sites belonging to a different EGEE[2] region than the Tier-1 itself.

Differing from the original tree structure of the WLCG, the current version of the CMS computing model implies that Tier-2 centres get the experiment data for further analysis not from one specific Tier-1 but from any Tier-1 holding the required dataset. MC generated data, however, will always be stored at one particular Tier-1. The computing models of the other LHC experiments imply a closer connection between the Tier-1 and Tier-2 sites, where the associated Tier-1 site delivers data to the Tier-2 sites and provides long-term storage for the MC data generated at the Tier-2 sites.

Table 1 shows the current CPU and storage resources of GridKa and the yearly upgrade milestones until 2009.

GridKa			
Alice, Atlas, CMS, LHCb			
Resources	CPU	disk	tape
2006	2020 kSI2k	640 TB	960 TB
2007	3080 kSI2k	1390 TB	1830 TB
2008	8300 kSI2k	3860 TB	4460 TB
2009	12780 kSI2k	5880 TB	8700 TB

Table 1: The GridKa resources 2006 and planned upgrades. The quoted numbers are the sum of LHC and non-LHC resources. The share of the 4 LHC experiments, however, will be >90% in 2008 and later.

Tables 2, 3 and Figure 1 give an overview over the (future) Tier-2 sites associated with GridKa, their resources and upgrade plans. Since the funding is unclear for many of the sites, the quoted numbers of resources have to be regarded as preliminary in most cases. The list is as of begin 2006. Additional sites may be connected to GridKa in the future.

CONNECTING TIER-2 SITES – TECHNICAL ASPECTS

This section describes briefly technical and organizational aspects of the cooperation of GridKa and its associated Tier-2 sites.
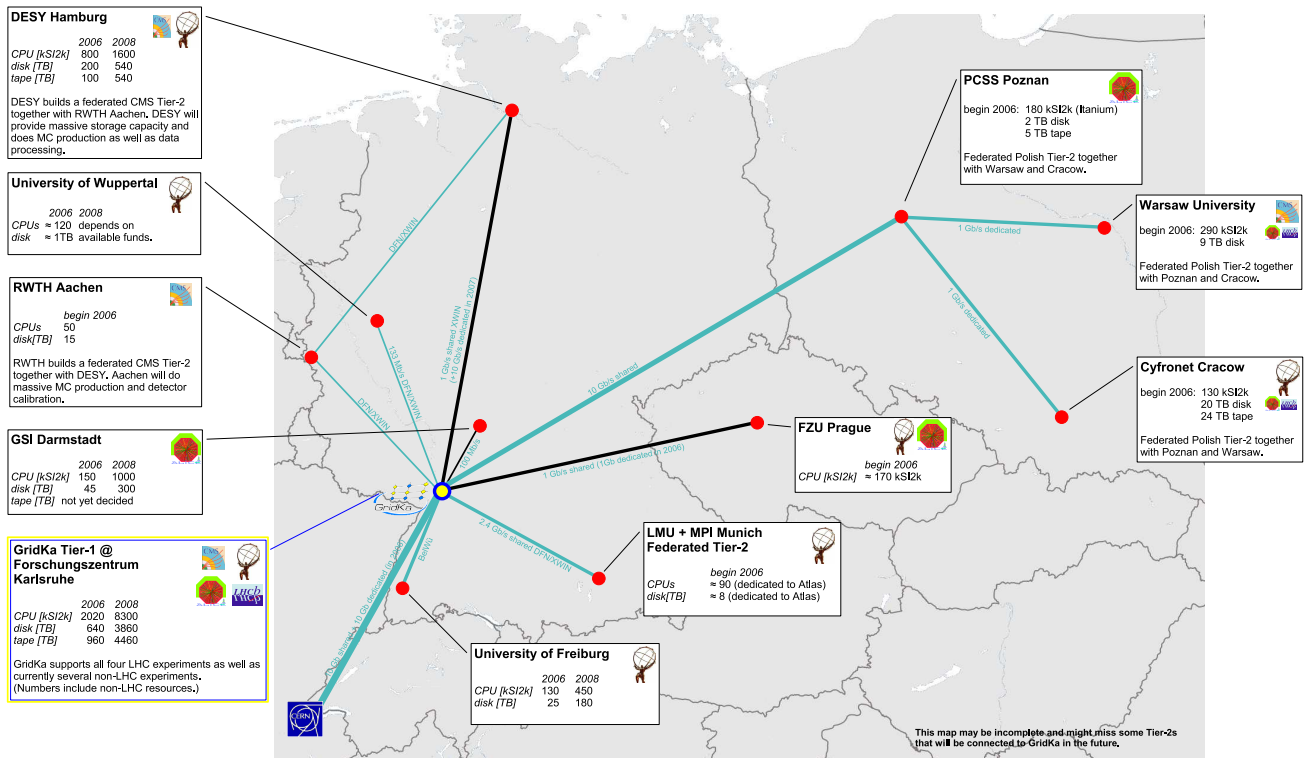
Figure 1: This map shows WLCG Tier-2 sites associated to GridKa as of begin 2006. The map is not complete. The CPU and storage resources as well as the network bandwidth to GridKa are indicated, though these numbers have to be regarded as preliminary.

Network

The wide area network (WAN) connection between GridKa and the Tier-2 sites has to provide the necessary bandwidth required by each individual site. This bandwidth is in the order of 1 to 10 Gb/s depending on the size of the Tier-2 centre and the experiments it supports. The expected network traffic volume between the associated sites varies in the different experiments computing models. For smaller Tier-2 sites, a (shared) connection using the standard internet infrastructure may be sufficient while for larger Tier-2 sites, such as DESY, a dedicated virtual private network (VPN) link is a necessity. The Tier-1 – Tier-1 transfer rates and especially the Tier-0 – Tier-1 transfers must not be influenced by the Tier-1 – Tier-2 network traffic. GridKa uses the DFN/XWIN (Deutsches Forschungsnetz)[3] national scientific network and Geant2[4], a pan-European data communication network to realize its WAN connections.

First data transfer tests have been performed with several of the associated Tier-2 sites. Figure 2 shows the results of such tests performed automatically in December 2005. The achievable transfer rates from GridKa to FZU Prague and to DESY have been measured hourly over several days using the srmcp[5] tool. The histograms show the daily average. The transfers were routed through a shared network infrastructure with 1 Gb/s maximum available bandwidth.
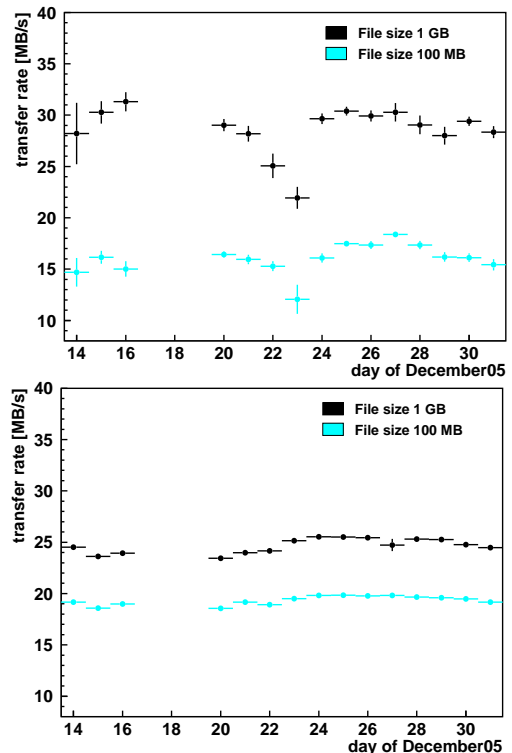

Figure 2: Daily averaged transfer rate of real file transfers between GridKa and DESY (top) and FZU Prague (bottom).

DESY Hamburg			
Atlas, CMS			
(federated CMS Tier-2 with RWTH Aachen)			
Resources	CPU	disk	tape
2006	800 kSI2k	200 TB	100 TB
2008	1600 kSI2k	540 TB	540 TB
2009	2100 kSI2k	870 TB	870 TB

RWTH Aachen			
CMS			
(federated Tier-2 with DESY Hamburg)			
Resources	CPU	disk	tape
2006	50	15 TB	-
2008	540	60 TB	-

GSI Darmstadt			
Alice			
Resources	CPU	disk	tape
2006	150 kSI2k	45 TB	-
2008	1000 kSI2k	300 TB	?

University of Freiburg			
Atlas			
Resources	CPU	disk	tape
2006	130 kSI2k	25 TB	-
2008	450 kSI2k	180 TB	-

LMU + MPI Munich			
Atlas			
Resources	CPUs	disk	tape
2006	90	8 TB	-

Table 2: Computing resources of German Tier-2 centres.

Security

The gridftp servers, especially those which are relevant for the data import from CERN, as well as the SRM[5] and other dCache hosts, need to be protected against attacks from the outside world. The specific security measures applied to a particular server or service depend on the connected network, i.e. public internet or VPN, and the data rates it has to serve. Where the utilisation of a firewall is not feasible due to performance reasons, the access to the particular servers will be limited to certain allowed hosts of the partner Tier-1 and Tier-2 sites by using the access control lists (ACL) at the relevant router at GridKa. Control and handshake connections, such as used by the SRM are not impaired by the firewall while gridftp data transfer rates would be affected when routed through a firewall which cannot handle the high gridftp data flows.

FZU Prague			
Atlas, Alice			
Resources	CPU	disk	tape
2006	≈170 kSI2k	?	-

Warsaw University			
CMS, Alice, LHCb			
(federated Tier-2 with Poznan and Cracow)			
Resources	CPU	disk	tape
2006	290 kSI2k	9 TB	-

PCSS Poznan			
Alice			
(federated Tier-2 with Warsaw and Cracow)			
Resources	CPU	disk	tape
2006	180 kSI2k (Itanium)	2 TB	9 TB

Cyfronet Cracow			
Atlas, Alice, LHCb			
(federated Tier-2 with Warsaw and Poznan)			
Resources	CPU	disk	tape
2006	130 kSI2k	20 TB	24 TB

Table 3: Computing resources of Tier-2 centres located outside the EGEE Germany/Switzerland region.

File Transfer Service

The File Transfer Service (FTS) is part of the gLite middleware used within the EGEE project and schedules and performs file transfers on unidirectional channels between sites. Each Tier-1 site hosts a FTS server, which runs the transfer channels to and from the Tier-2 sites. Transfer channels between Tier-1 sites are also hosted on the FTS server of the destination site. Each channel has its set of parameters, such as the number of simultaneous file transfers, number of streams and maximum available bandwidth. The available bandwidth can be shared between several virtual organisations (VOs) and the individual VO shares can be adjusted by the channel administrators. The authorization system of FTS forsees different roles. "Normal" users can manage their own jobs and may be able to submit new jobs but cannot change channel parameters. VO managers can get information on the different channels and can manipulate all transfer jobs belonging to their VO. Channel administrators can change parameters of a channel and set the channel active or inactive. A Tier-2 site administrator should be channel administrator of the transfer channel from the Tier-1 to his Tier-2 site and thus be able to control the data flux to his site.

Storage System

GridKa uses a dCache [7] type storage element. The system is currently setup in such a way, that access to the data is possible from the (internal) worker nodes as well as external Tier-1 and Tier-2 sites. The data stream from CERN to the storage system must not be interfered by other data accesses. This is guaranteed by using separate "pool nodes" (dCache storage nodes) for the different external and the internal clients. GridKa uses internal read- and write–pools, external read- and write–pools as well as special write–pools for the CERN–GridKa data stream. Figure 3 gives an overview of the GridKa dCache setup. The GridKa dCache system is described in more detail in [8].
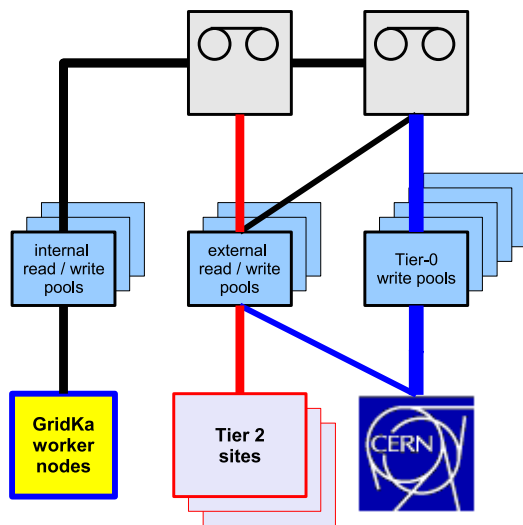

Figure 3: A simplified view of the GridKa dCache setup.

Site Certification

Within the context of the EGEE project, the German/Swiss (DECH) regional operations centre (ROC) is managed by GridKa. Among the ROC managers responsibilities are the testing of the middleware installations of new Tier-2 sites and the certification of the positively tested centres as EGEE resource centres. The complete certification procedure is well described in [9].

Tier-1 – Tier-2 Collaboration

A close cooperation between the sites is required in order to guarantee trouble-free data transfers between the Tier-1 and its associated Tier-2 sites. Changes in the configuration of the storage system and networking of a site have to be announced to the other sites immediately to allow for necessary changes of routing and firewall setups. The usage of a task and bug tracking system would certainly be advantageous. Here, the GGUS system[10] (Global Grid User Support) could be the suitable platform.

SUMMARY

Within the WLCG, the Tier-1 centres has important responsibilities. It has to store the raw experiment data coming from CERN with very high reliability. The CPU resources will mainly be used to process the raw data to new data formats used for analysis. Associated Tier-2 sites send their MC data to the Tier-1 for long-term storage and request processed data sets. The network and storage system setup of GridKa allows for performing these tasks without vitiating each other. A close collaboration with the Tier-0 and the Tier-2 sites is necessary to provide the high quality services required by the LCG experiments and to adapt to the probably changing requirements in the long term.

REFERENCES

[1] The Worldwide LHC Computing Grid (WLCG)
http://lcg.web.cern.ch/lcg

[2] The "Enabling Grids for E-SciencE" project (EGEE)
http://www.eu-egee.org

[3] Deutsches Forschungsnetz – Germany's National Research and Education Network, http://www.dfn.de/content

[4] Geant2
http://www.geant2.net

[5] Storage Resource Manager, http://www-isd.fnal.gov/srm

[6] File Transfer Service (FTS)
http://egee-jra1-dm.web.cern.ch/egee-jra1-dm/FTS/

[7] dCache
http:/www.dcache.org

[8] D. Ressmann, S. Halstenberg, J. van Wezel: "Using TSM to create a high-performance tape connection", CHEP 2006

[9] Site Registration Policy and Procedure
http://edms.cern.ch/document/503198

[10] GGUS - Global Grid User support
http://www.ggus.org