Integrating a heterogeneous and shared Linux cluster into grids

V. Büge,[1,2] U. Felzmann,[1] C. Jung,[1,2] U. Kerzel,[1]
M. Kreps,[1] G. Quast,[1] A. Vest[1]

1 UNIVERSITÄT KARLSRUHE

2 IWR Institut für Wissenschaftliches Rechnen
Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft

XV International Conference on Computing in High Energy and Nuclear Physics

February 13 – 17, 2006

Tata Institute of Fundamental Research,
Mumbai, India

Outline

- Why do university groups need grid computing?
- Computing environments at universities
  - Representative example: IEKP cluster at Karlsruhe
  - Peculiarities of a typical university cluster
- Grid middleware requirements and site specific grid services
- The IEKP LCG site
- Conclusion and Outlook

Why do university groups need grid computing?

- Current and future HEP experiments at Tevatron or LHC: huge data production rates and event sizes

  → Worldwide distributed datasets already now:

  - Simulated data in the LHC experiments, O(100 TB)
  - Real data in CDF, D0, H1, ZEUS, Babar, etc., O(1 PB)
- Processing power widely available in associated institutes
  - Collaborating groups cope with these challenges using grid tools
  - Opportunistic/shared use of the resources between local users and grid users
- Benefits of integrating an institute's cluster into grids:
  - Minimisation of idle times
  - Interception of peak loads
  - Shared data storage
  - Shared deployment effort of common services

Computing environment at universities

A typical university computer cluster has to cope with diverse challenges:

- Heterogeneous structure in:
  hardware, software, funding and ownership
- Support of multiple groups with different applications and sometimes conflicting interests
- Infrastructural facilities have grown in the course of time
  → characteristic history and resulting inhomogeneities
- Embedded in structures imposed by institute, faculty and university

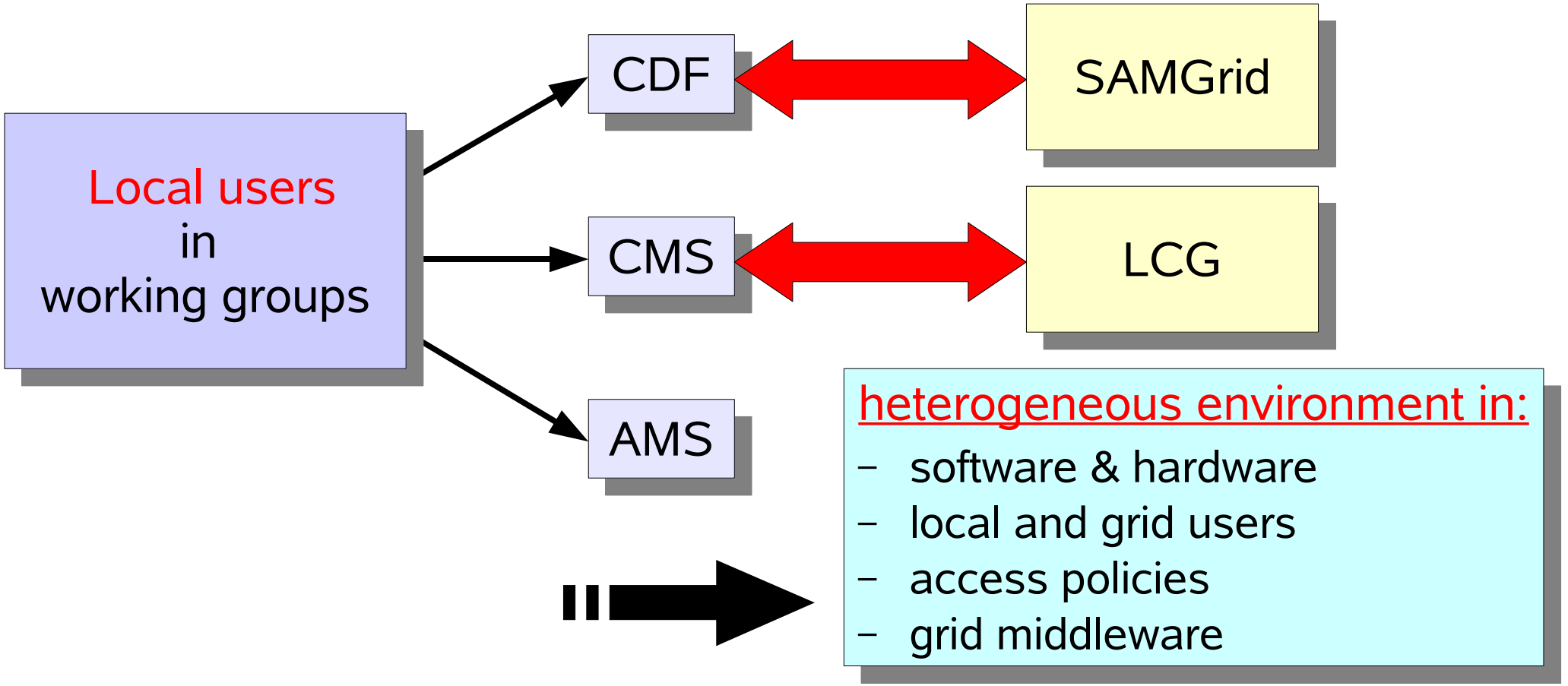
➡ Integration into existing grids not easy at all!

- Idea of sharing resources still not present in all minds

Example: IEKP cluster at Karlsruhe

Representative example: IEKP (Institut für Exp. Kernphysik) at the University of Karlsruhe integrated in:

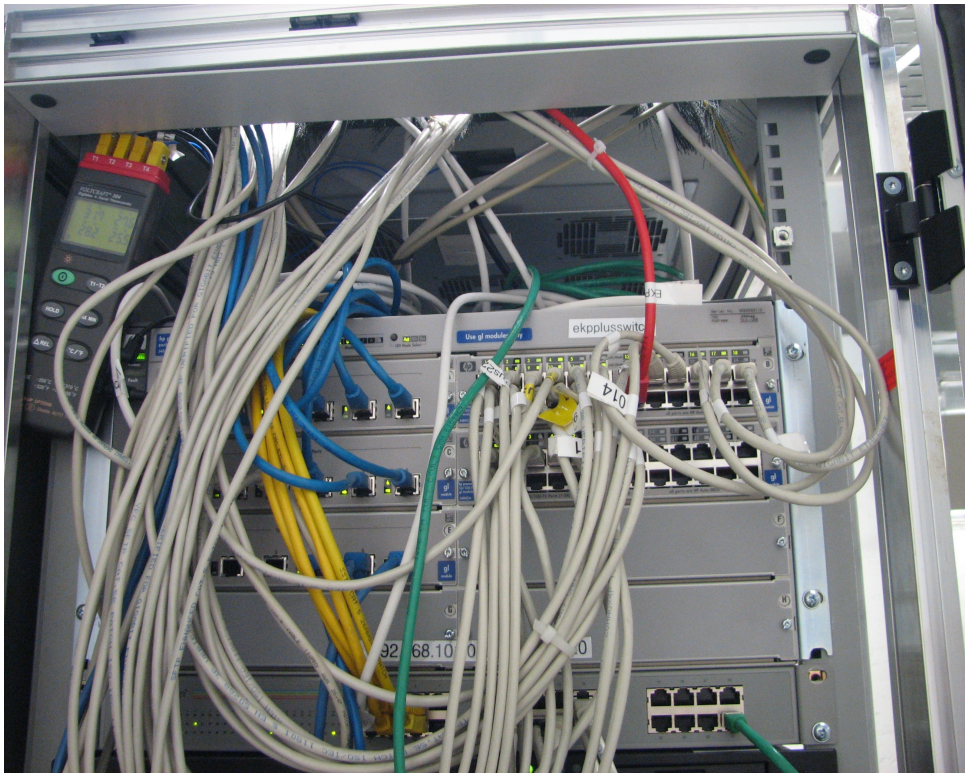
- SAMGrid (Sequential Access via Metadata Grid) for CDF and
- LCG (LHC Computing Grid) for CMS



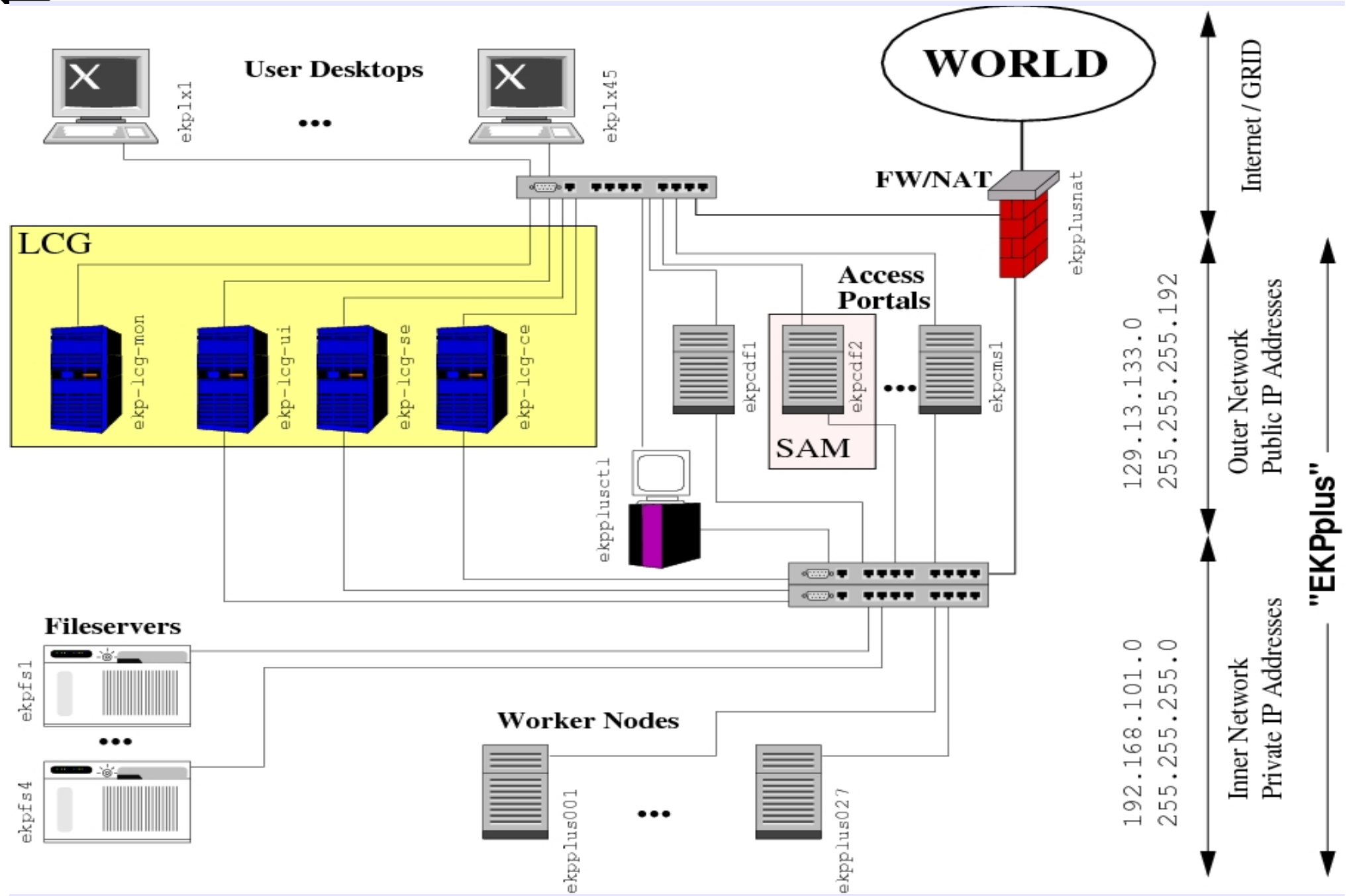

Map
Satellite
Hybrid
GridKa/FZK
(Tier-1)
IEKP
(Tier-3)

Example: IEKP at Karlsruhe

- IEKP Linux cluster component specifications:
  - One or more portal machines for each experiment (3 for CDF, 1 for CMS, 1 for AMS)
  - 5 file servers, ~ 15 – 20 TB disk space
  - 27 Computing Nodes, 36 CPU's
- Linux cluster independent of desktop cluster



Architecture of the IEKP Linux cluster



Peculiarities of a typical university cluster

- Network architecture:
  - Inner network:
    computing nodes, file servers, cluster control machine
  - Outer network:
    publicly accessible portals
    - **development of analysis software**
    - connection to inner network
      → access to file servers
      → usage of Worker Nodes via the local batch system
- Network protocols and services:
  - User accounts exported by the cluster control machine to all nodes via Network Information Service (NIS)
  - File/root systems exported via Network File System (NFS)
  - Other supported protocols: GSIFTP, SRM

Peculiarities of a typical university cluster

- Local batch system: PBS/Torque (Open Source)
  - Scheduler: MAUI
    fair share principle (group and user fair share)
  - Batch queues
- Firewall
  - Grid components behind the firewall of the institute
  - Allow external access to grid services:
    some ports of the firewall have to be opened
  - Internal campus net: protected by the university's computing department (switched off for IEKP cluster)
- Desktop cluster
  - User wokstation and access point to the portal machines
  - IEKP Linux cluster:
    connected to the desktop network by a 1 GBit connection

Operating systems

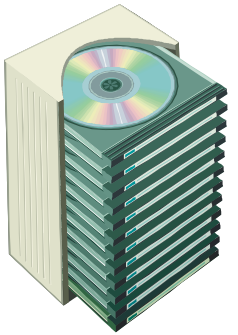
- Software on portal machines and Worker Nodes (WNs): experiment dependent
- Operating system on all machines: Linux,
  but the flavour is not identical on all components
  (due to experiment specific extensions/modifications)
  - CDF portals: Linux distribution based on Fermi RedHat 7.3
  - CMS portals: Scientific Linux CERN 3
  - WNs: Scientific Linux CERN 3
    (the only OS under which all AMS, CDF, CMS software and grid software runs, at least for the moment)
- No major problems occurred running different Linux distributions on the same cluster!
- WNs: 32-bit operating system;
  upgrade to a 64-bit operating system foreseen

Grid middleware requirements

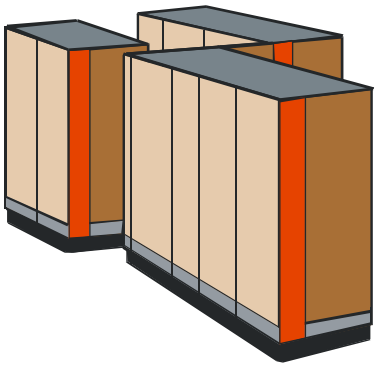
- Flexibility:
  Installation procedure and setup of the grid middleware modifiable according to the local conditions
- Interoperability:
  Compatibility with other grid middlewares
- Dynamic:
  Possibility to add or remove resources during the running grid service
- Encapsulation:
  Shielding of experiment and analysis software from changes in the underlying grid environment
- Level of abstraction:
  Access to computing and storage resources must be independent of their physical location and local setup

Site specific grid services (SAMGrid)

- SAMGrid station
  - Only one dedicated machine needed per cluster (portal machine)
    - User interaction
    - Mass storage
    - File import / export / delivery to analysis programs
    - File exchange via GSIFTP
    - Activity is written to central database



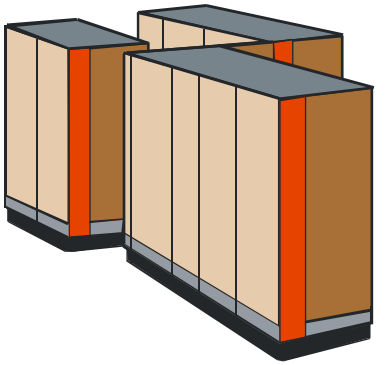
(U. Kerzel *et al.*, CHEP06, Id153)

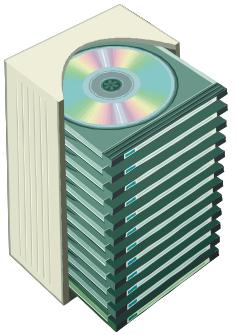


Site specific grid services (LCG)

Layers of abstraction, local peculiarities irrelevant

- Computing Element (CE)
  - Gateway to local computing resources, i.e. Worker Nodes via local batch system
  - Globus Gatekeeper
- Storage Element (SE)
  - Gateway to local storage (disk, tape)
  - Globus GridFTP server, SRM interface
- Monitoring Box (MON)
- User Interface (UI)
  - User's access point to the grid and portal for (grid) file access
  - Client programs using grid services
  - Could also be a desktop machine





The IEKP LCG site

- IEKP site: Tier-2/3 prototype centre within LCG
- Dedicated configuration → IEKP offers full grid functionality:
  - Grid based physics analyses
  - Software installation for grid and local users
  - Data storage
- No Virtual Organisation Management Service (VOMS) yet but: need for a prioritisation of certain user groups!
- Present workaround: treatment of grid users using the IEKP capacities depending on their affiliation organised by
  - Mapping to different accounts and user groups
  - Configuration of the corresponding batch queues, e.g. different queue priorities managed by fair share targets:
    - Local CMS users (cms) → local accounts
    - German CMS users (dcms) → generic mapping accounts
    - CMS Collaboration (cmsgrid) → generic mapping accounts

Conclusion & Outlook

- Computer clusters at universities:
  often shared and situated in a heterogeneous environment
- IEKP cluster successfully integrated in SAMGrid and LCG without compromising local user groups
  - Virtualisation of the LCG components foreseen
  - IEKP-KA/2006-3: "Integrating the IEKP Linux cluster as a Tier-2/3 prototype centre into the LHC Computing Grid"
- <u>Many thanks to:</u>
  - the German Bundesministerium für Bildung und Forschung BMBF
    for financial support

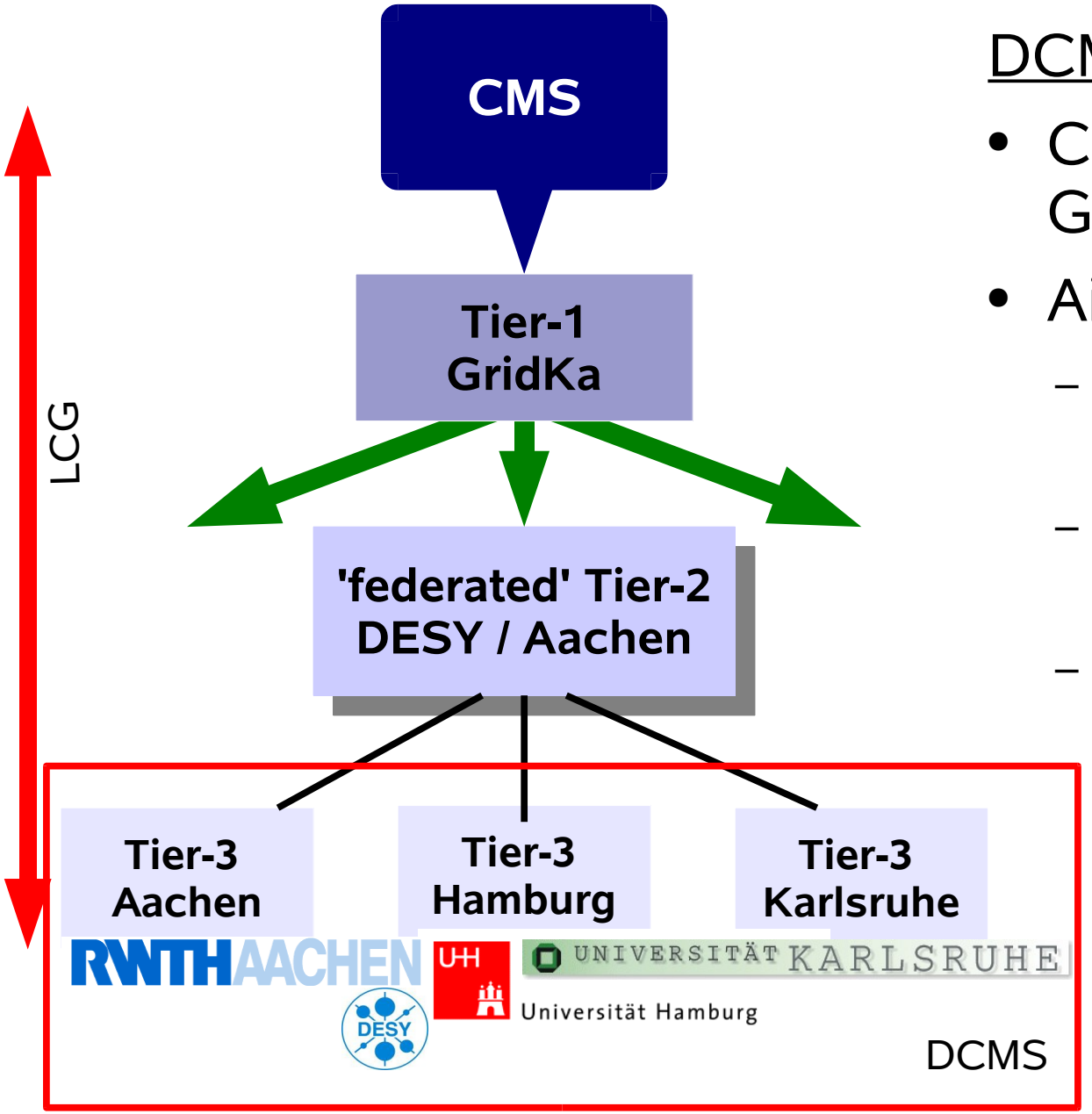
  - the Universitäts-Rechenzentrum Karlsruhe for technical support
  - the IEKP system administrators
    Y. Kemp, M. Milnik, C. Sander and P. Schemitz.
  - the dcms LCG administrators
    A. Gellrich, M. de Riese, M. Kirsch and A. Nowack

Conclusion & Outlook

A successful and beneficial integration of existing university clusters into computing grids is possible, but we have a wishlist:

- Precondition:
  Motivate university groups to share their free computing resources within grid environments
- A must:
  Mechanisms for priorisation, accounting and charge-back ("billing")
- Operating system independence of experiment specific and grid software, in particular the LCG middleware
- Ease the LCG installation on existing computing environments:
  lightweight and non invasive installation procedure needed,
  in particular concerning Worker Nodes
- Virtualisation of LCG environment/hosts on a single powerful server/node (improves utilisation of resources and security aspects)
- Virtualisation of Worker Nodes may cope with significant differences of the Linux flavours required by certain user groups

Present CMS grid structure in Germany



DCMS:

- Collaboration of the German CMS institutes
- Aims:
  - Exchange of experience in analyses (on the grid)
  - Sharing of available resources
  - Prioritisation of dcms users at the DCMS sites

VO dcms will hopefully become obsolete with VOMS