Die Tier2-Site DESY-ZN

die langweiligen technischen Details

Andreas Haupt
<andreas.haupt@desy.de>

Agenda

- Begriffsklärung
- Netzaufbau
- Softwarebereitstellung
- Pakete und deren Funktion
- Vorstellung einzelner Dienste
- Weitere Informationsquellen

Abkürzungen

- LCG LHC Computing Grid
- UI User Interface
- CE Computing Element
- RB Resource Broker
- WN Worker Node
- SE Storage Element
- SRM Storage Resource Manager
- GRIS Grid Resource Information Server
- GIIS Grid Information Index System
- BDII Berkeley Database Information Index
- R-GMA Relational Grid Monitoring Architecture

Was ist ein ... (1)

- User Interface
  - Desktop oder Workstation mit installierter Middleware (gLite)
- Worker Node
  - Batch-Farmknoten mit installierter Middleware
- Computing Element
  - verwaltet und überwacht Grid-Jobs (Annahme vom RB, Übergabe an das lokale Batchsystem, ...)

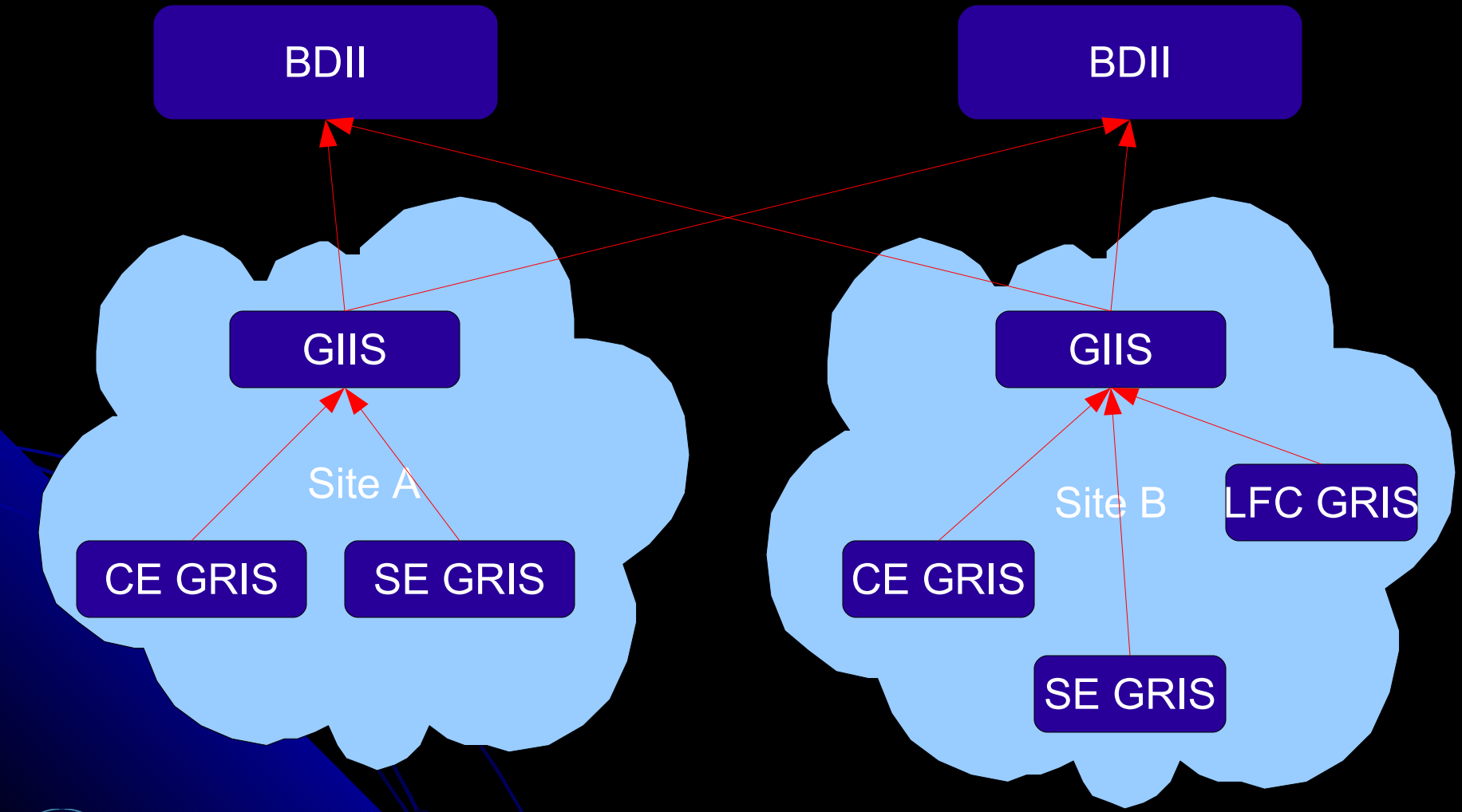
Was ist ein ... (2)

- Resource Broker
  - nimmt Jobs vom UI entgegen und leitet sie an ein passendes CE weiter
- Storage Element
  - ein SRM mit dem Dateien über Gridprotokolle verwaltet werden können
- GIIS (oder Site-BDII)
  - sammelt die LDAP-Daten der GRISes einer Site und stellt sie gesammelt (wiederum per LDAP) den BDIIs zur Abfrage zur Verfügung
  - TCP Port: 2170

Was ist ein ... (3)

- GRIS
  - LDAP-Server, der Daten über den aktuellen Zustand der Dienste auf einem Grid-Host zur Verfügung stellt
    - GlueSchema
  - TCP Port: 2135
- BDII
  - LDAP-Server, der die GIISes aller zertifizierten Sites abfragt und die Ergebnisse zusammenstellt
  - TCP Port: 2170

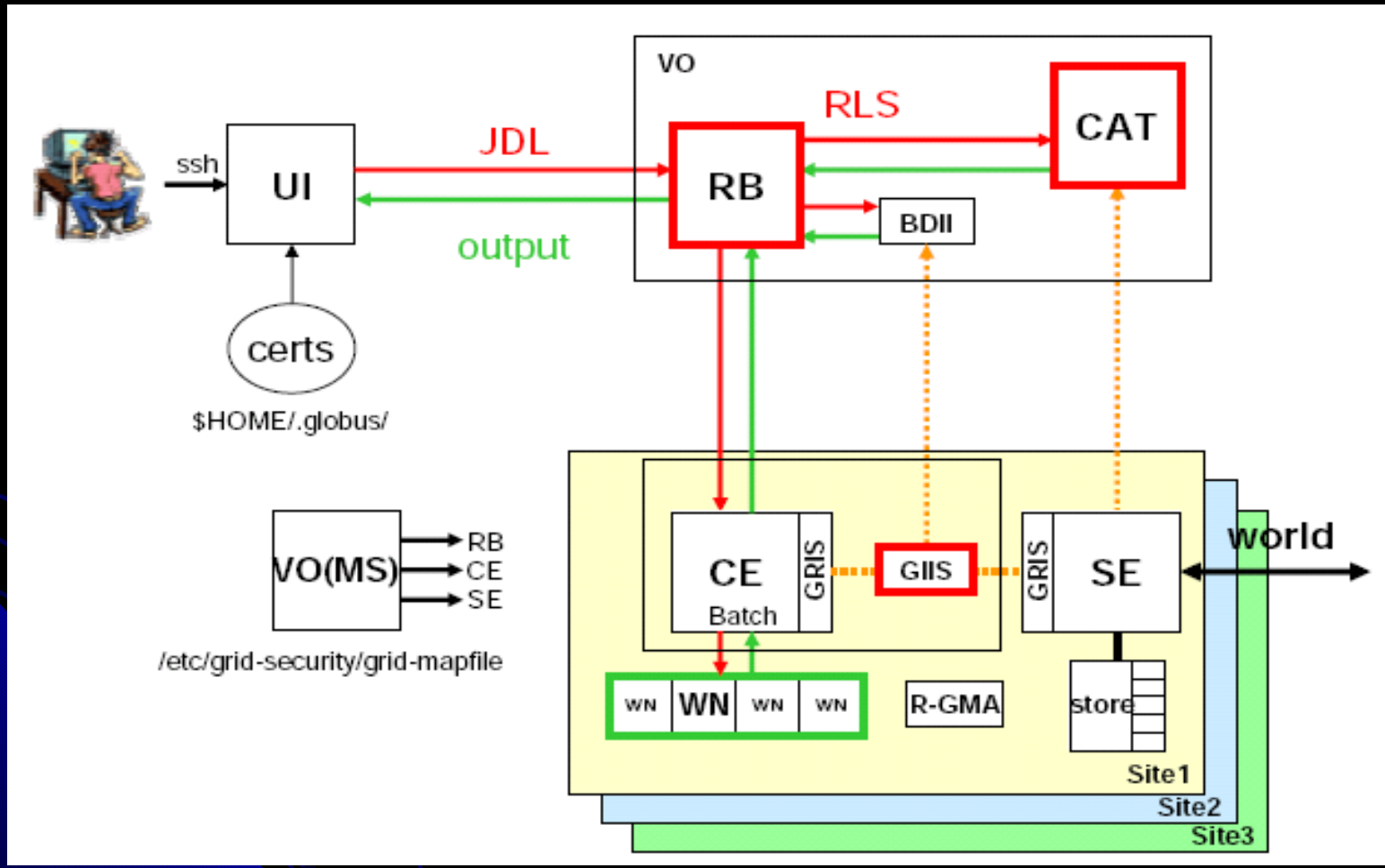
Das globale Informationssystem



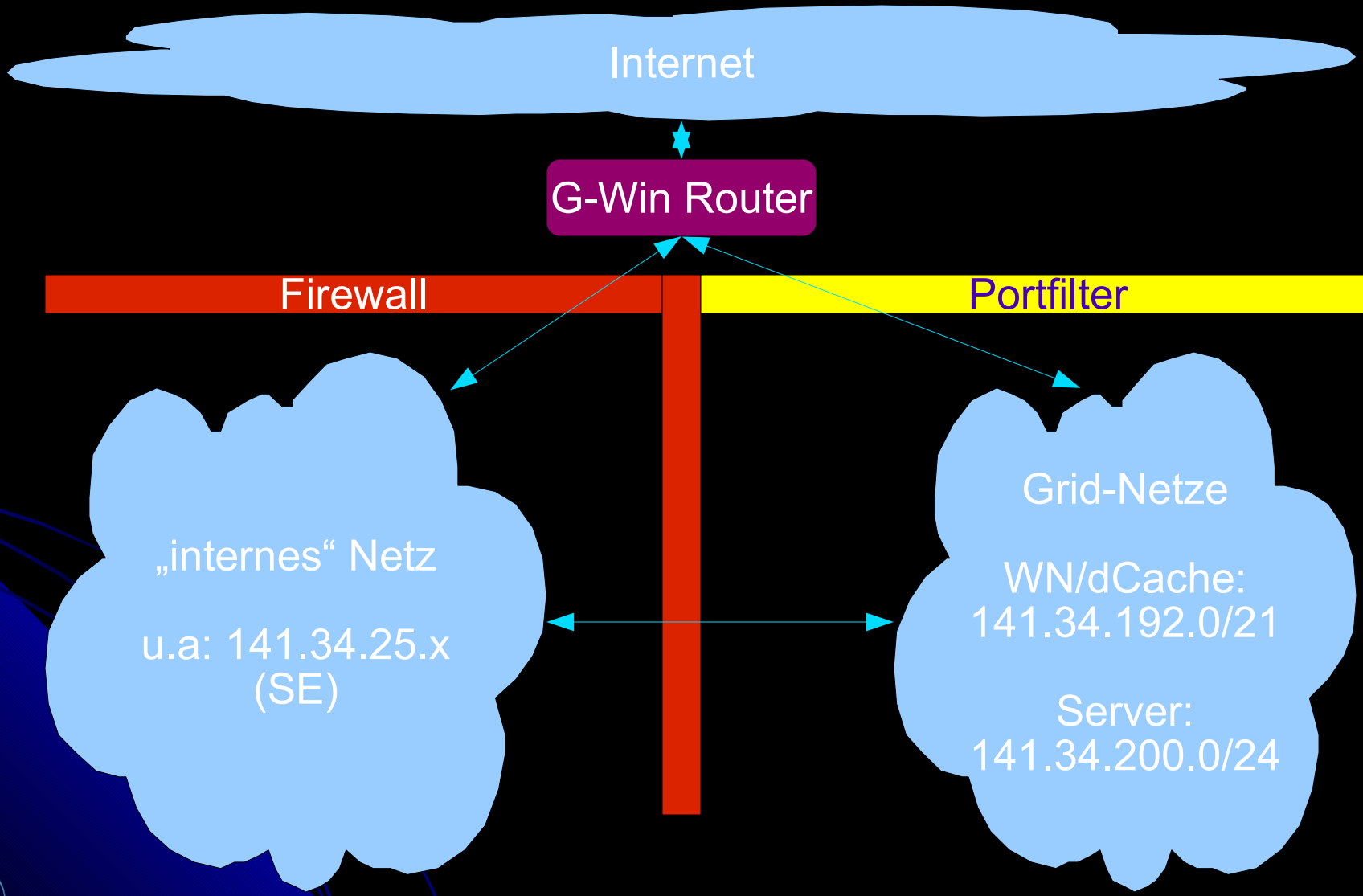
Was ist ein ... (4)

- R-GMA
  - Webservice, der eine Art verteilte Datenbank zur Verfügung stellt
  - Abfragen, Änderungen per SQL möglich
  - wird auch benutzt, die Accountingdaten aller Sites zu sammeln (APEL)
  - läuft auf der MON-Box einer Site
  - leider nicht sonderlich stabil und sehr träge

Der „Grid-Jobfluss“



Netzaufbau


Internet
G-Win Router
Firewall
Portfilter
„internes“ Netz
u.a: 141.34.25.x
(SE)
Grid-Netze
WN/dCache:
141.34.192.0/21
Server:
141.34.200.0/24

Softwarebereitstellung

- Mirror von SL sowie gLite auf a
  - /nfs1/a/SL/gLite/3.0 – Middleware
  - /nfs1/a/SL/gLite/LCG-CA - Zertifikate
- bereitgestellt über den Apache auf a
- eigene Pakete:
  - $TIER2_PATH/RPMS/<arch>

Installation (1)

- /project/linux/Tier2/SL-<Version>-gLite-<Version>
  - /project/linux/Tier2/SL-308-gLite-3.0/
- SL3 Kickstart
  - eine einheitliche! Konfigurationsdatei für alle Rechner einer SL-Version
- viel Logik im Preinstallationsskript
  - Partitionierung mit Erhalt einer Servicepartition
  - Ermittlung des Hostnamens / IP-Konfiguration
  - etc

Installation (2)

- Postinstallationsskript übernimmt Basiskonfiguration
  - Syslog, Mail, ...
  - Installation des T2_base Pakets
    - yum- und yumsel-Konfiguration
    - einige Skripte, die überall vorhanden sein sollen
    - Ermittlung des Typs (CE, WN, GIIS, ...) anhand des Hostnamens
  - /etc/yumsel.d/99_desy-zn-<Typ>.ys schreiben
    - enthält Namen des RPMs, das den Rechner konfiguriert (T2_DESY-ZN-<Typ>)

Pakete und deren Funktion (1)

- T2_base
  - Basispaket auf allen Tier2-Rechner
- T2_DESY-ZN-<Typ>
  - Konfiguriert den entsprechenden Hosttyp und zieht mit seinen Abhängigkeiten die Installation der nötigen Pakete nach sich
- T2_siteinfo
  - enthält /etc/lcg/site-info und läßt im Postinstall-Skript yaim laufen

Pakete und deren Funktion (2)

- T2_etc-files
  - enthält Vorlagen von /etc/passwd, /etc/group, /etc/shadow, etc. der Poolaccounts, die im Postinstall-Skript in die „echten“ Dateien eingepflegt werden
  - legt die Heimatverzeichnisse der Poolaccounts an
- T2_ssh-files
  - enthält ssh_known_hosts2 Datei
  - konfiguriert paßwortlosen Zugang der Poolaccounts von WNs auf CE

Pakete und deren Funktion (3)

- T2_vomscerts
  - enthält VOMS-Zertifikate der nicht-LHC VOs (DESY VOs, DECH)
- T2_fixup_javapath
  - stellt einheitlichen Pfad zum Java RE, SDK unter /usr/java/default bereit
  - Link auf die gerade aktuell installierte Version

Pakete und deren Funktion (4)

- T2_ntp
  - konfiguriert Zeitservice
- T2_console
  - konfiguriert die serielle Konsole

yaim

- Installations- und Konfigurationstool der gLite-Middleware
- csh-Skript !!!
  - einzelne Dateien als Konfigurationsplugins (config_<plugin>)
  - relativ simpel mit 'perl -pi -e ...' editierbar...
- node-info.def enthält Pluginnamen für einen bestimmten Hosttypen
- Aufruf: yaim <siteinfo> <Typ>
  - z.B.: yaim /etc/lcg/site-info CE

CE (1)

- Services:
  - Globus-Gatekeeper (nimmt die Jobs vom RB entgegen)
  - GRIS: LDAP-Service (aktueller Status des CE und der Jobs)
  - GridFTP: Proxyzertifikat, etc. vom RB kopieren
- Besonderheiten:
  - Alle Poolaccounts müssen sich von den WN auf dem CE paßwortlos einloggen können! (Braucht Torque zum „Stage in“ des Proxyzertifikats)

CE (2)

- LCMAPS:
  - Mapping Zertifikats-DN -> Poolaccount
  - /etc/grid-security/grid-mapfile
    - Mapping Zertifikats-DN - Accountpool
  - /etc/grid-security/gridmapdir
    - Hardlinks Poolaccountname – Zertifikats-DN
    - speichert einmal vorgenommene Mappings zur Wiederverwendung ab
  - VOMS:
    - /opt/edg/etc/lcmaps/gridmapfile

CE (3)

- GRIS:
  - Vorlagen in ldif-Dateien
  - Plugins:
    - Modifizieren einzelne Verzeichnisse im LDAP-Baum
  - Provider:
    - Fügen neue Verzeichnisse dem LDAP-Baum hinzu
  - Output von Plugins, Provider und Vorlage wird zusammengestellt und ergibt LDAP-Output
  - liegen unter /opt/lcg/var/gip

Torque / Maui (1)

- Laufen auf lcg-bm
- Torque Konfiguration:
  - qmgr -c 'p s'
  - Dateien unter /var/spool/pbs
  - Konfigurierte Knoten
    - „pbsnodes -l“
    - /var/spool/pbs/server_priv/nodes
  - Accountingdaten:
    - /var/spool/pbs/server_priv/accounting

Torque / Maui (2)

- Maui Konfiguration
  - /var/spool/maui/maui.cfg
- Kommandos:
  - showq (-r)
  - diagnose (-p, -f)
  - showconfig
- http://dvinfo.ifh.de/Grid/Admin_Info

SE (1)

- globe-door.ifh.de
- SRM stellt den externen Zugriff auf den dCache zur Verfügung (Webservice)
  - Kein Datentransfer!
    - das machen GridFTP, (GSI-)dCap
  - Authentisierung über GSI
  - Operationen:
    - „Stagen“ von Dateien vom Band
    - Anlegen / Löschen von Verzeichnissen
    - TURL für eine SURL und ein bestimmtes Protokoll ermitteln

SE (2)

- ist Teil des dCache-Pakets
- Konfiguration in:
  - /opt/d-cache/etc/node-config
  - /opt/d-cache/config/dCacheSetup
  - /opt/d-cache/etc/dcache.kpwd
- benutzt Postgres-Datenbank (Name: dcache) zum Loggen:
  - getfilerequests(_b)
  - putfilerequests(_b)
  - ...

SE (3)

- globe-door hat auch GSI-dCap Tür
- GRIS mit dCache-eigenem Infoprovider
  - das übliche Skript reicht nur den Output dieses Providers durch
  - verlangt etwas eigenwillige Poolmanager-Konfiguration:
    - viele Poolgruppen – heißen wie die VOs
    - Pools werden den Poolgruppen zugeordnet

dCache Poolknoten

- im selben Subnetz wie die WNs
  - soll hohen Gesamtdurchsatz garantieren
- zwei Dienste:
  - dcache-pool – startet den Pool-Service
  - dcache-core – startet den GridFTP-Service
    - jeder Poolknoten besitzt eine GridFTP-Tür – soll zu einer besseren Skalierung bei vielen gleichzeitigen Transfers beitragen
- Gleiche Konfigurationsdateien wie SRM
  - /opt/d-cache/[etc|config]/...

MON-Box

- Client Zugriff auf R-GMA über Tomcat-Webservice (Port 8443)
- MON-Boxen unterhalten sich über Port 8088
- APEL-Datenbank:
  - lokales MySQL
  - Accountingdaten der CEs werden hier aufbereitet
  - publiziert die gesammelten Accountingdaten jede Nacht

Weitere Informationen

- http://dvinfo.ifh.de/Grid/Admin_Info
- http://dvinfo.ifh.de/Grid/Information_Sources
- http://grid.desy.de/