DESY/Aachen Tier2 Site Report

Michael Ernst

DESY

With contributions from

Thomas Kress

RWTH Aachen

FZK T1-T2 Workshop October 2005

Building a Tier 2 Center at DESY for ATLAS and CMS

- **DESY offers to run an average Tier 2 center each for ATLAS and CMS**
  - Long term commitment (~20 years)
  - Three year project for ramp up
  - Will become part of standard computer center services

- **The Team:**
  - Michael Ernst (Applications, User, Service and Data Challenges)
  - Patrick Fuhrmann (Mass Storage Software)
  - Martin Gasthuber (Fabrics)
  - Andreas Gellrich (Grid)
  - Volker Gülzow (Project Leader)
  - Knut Woller (HW-Technology)

Grid Services at DESY

- OS (common to all services): SL 3.04
- Quattor (OS for all nodes; complete installation for WNs)
- Yaim (for all service nodes)
- LCG-2_6_0
- Central VO Services
  - VO Server (LDAP)   [grid-vo.desy.de]
  - VOMS (being investigated)   [grid-voms.desy.de]
  - Replica Location Service (RLS)   [grid-cat.desy.de]
- Distributed VO Services
  - Resource Broker (RB)   [grid-rb.desy.de]
  - Information Index (BDII)   [grid-bdii.desy.de]
  - Proxy (PXY)   [grid-pxy.desy.de]
- Site Services
  - GIIS: `ldap://grid-giis.desy.de:2170/mds-vo-name=DESY-HH,o=grid`
  - 2 CEs: 100 CPUs now, will be >200 CPUs by end 2005
  - SE: dCache based w/ access to entire DESY data repository
    - Storage: disk 15 TB (+ 40 TB by end 2005), tape 1 PB media (4 PB capacity)

Connectivity & Infrastructure

- DESY will have a 1 Gbps (raw) DFN/XWIN connection in 2006
    - Initially contracted: 300 Mbps
    - To be upgraded to 600 Mbps if demanded
- VPN connection to GridKa (p2p, likely 10 Gbps in 2007) in addition to 1 Gbps shared XWIN
- Additional 300 sm floor space will be ready in spring 2006
    - Power, UPS and Cooling is already prepared
- Share monitoring as has been set up for HERA-Computing

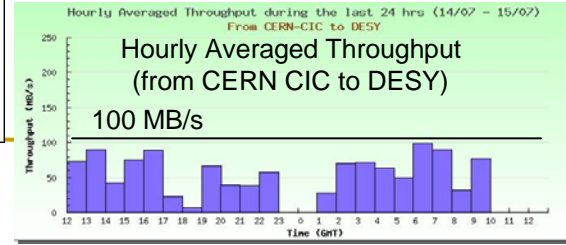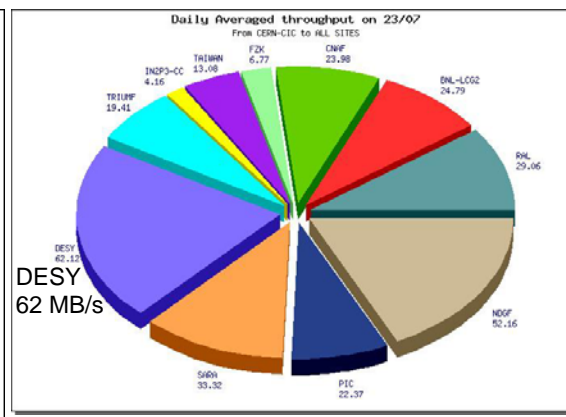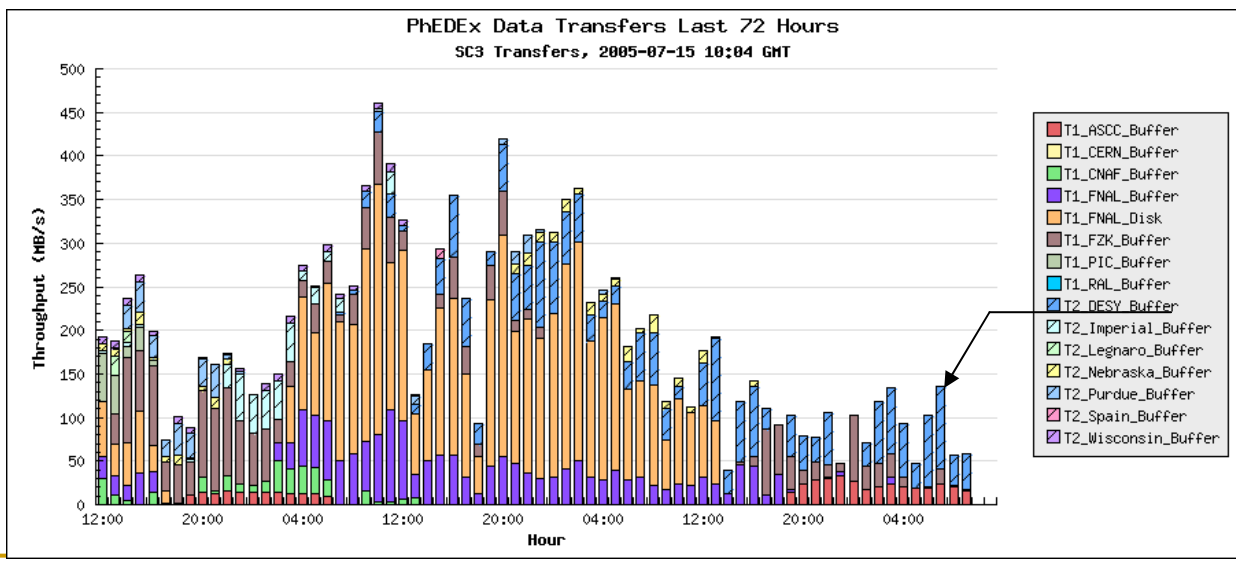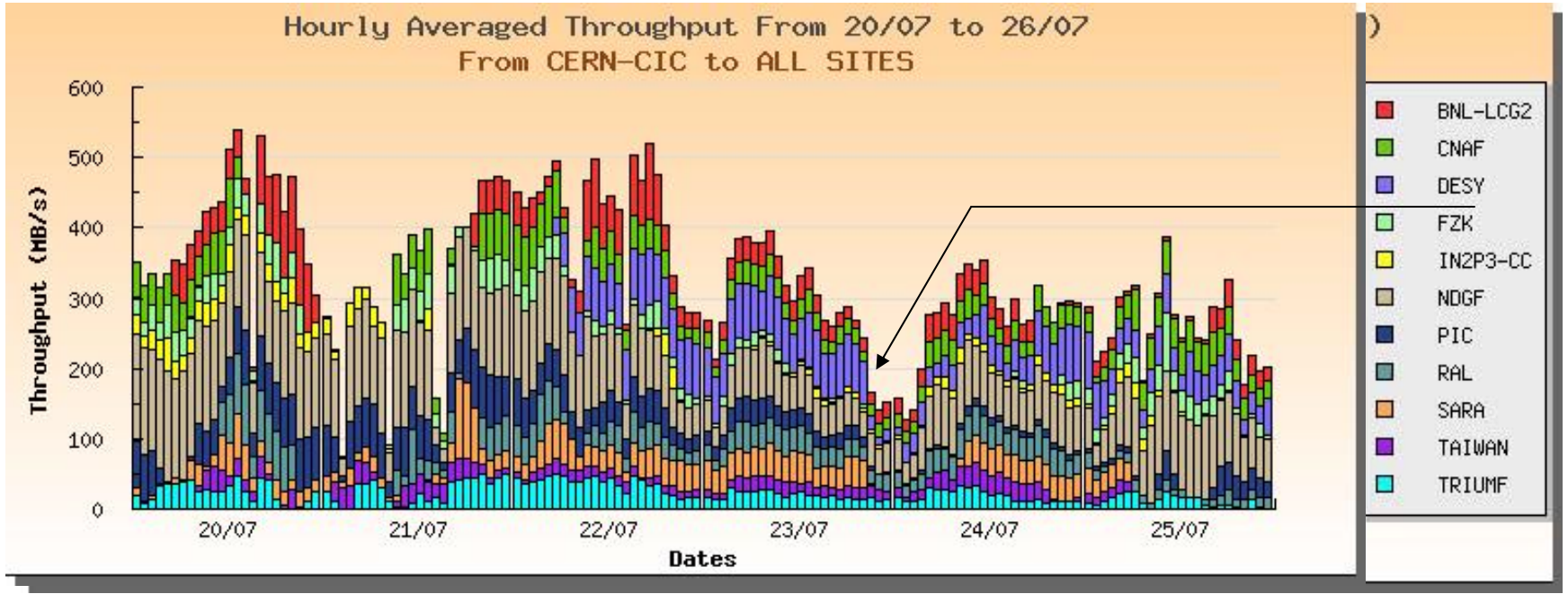# DESY HH: Proposed Hardware Resources *

(A=Atlas, C=CMS)

	2005	2006	2007	2008	2009
CPU [kSI2k]	100 A 100 C	400 A 400 C	700 A 700 C	700 A 900 C	900 A 1200 C
Disk [TB]	15 A 15 C	100 A 100 C	340 A 200 C	340 A 200 C	570 A 300 C
Tape [TB] (?)	10 A 10 C	50 A 50 C	200 A 100 C	340 A 200 C	570 A 300 C

* Ramp up subject to available funding

Information on DESY Tier-2 SC3 status

	Current Situation			
Core Components	**Status**	**Comments**	**July (throughput)**	**September (service)**
Network link	RUNNING	1 Gbit/sec shared with production activities	1 Gbit/sec shared	1 Gbit/sec shared
SRM v1.1	RUNNING	SRM/dCache (currently 15TB, 55 TB in Dec.)	SRM/dCache	SRM/dCache
Mass Storage	RUNNING	STK Silos w/ 9940B Drives	Production Service, 10TB Tape Media	Production Service, 10TB Tape Media
LFC	RUNNING	Version 1.3.4, MySQL	Production Service	To be used as local catalogues for the experiments.
FTS	Installed	Client	Production Service	Planned for ATLAS T1-T2 transfer
VO Components	**Status**	**Comments**	**July (throughput)**	**September (service)**
CMS Phedex	RUNNING	Being tuned with SRM	Production Service	Production Service
CMS PubDB	RUNNING	Catalog used for DS auto discovery	Pre-Production	Production Service
CMS Catalog	Preparing for Installation	LFC with POOL 2.1.0 being evaluated by CMS	POOL mysql	POOL LFC

http://grid.desy.de/sc/sc3/DESY-SC-status.htm

Hourly Averaged Throughput From 20/07 to 26/07
From CERN-CIC to ALL SITES

PhEDEx Data Transfers Last 72 Hours
SC3 Transfers, 2005-07-15 10:04 GMT

Daily Averaged throughput on 23/07
From CERN-CIC to ALL SITES

DESY
62 MB/s

Hourly Averaged Throughput
(from CERN CIC to DESY)

100 MB/s

Hourly Averaged Throughput during the last 24 hrs (14/07 - 15/07)
From CERN-CIC to DESY

Per file transfer performance with 1 to 30 files per srmcp request
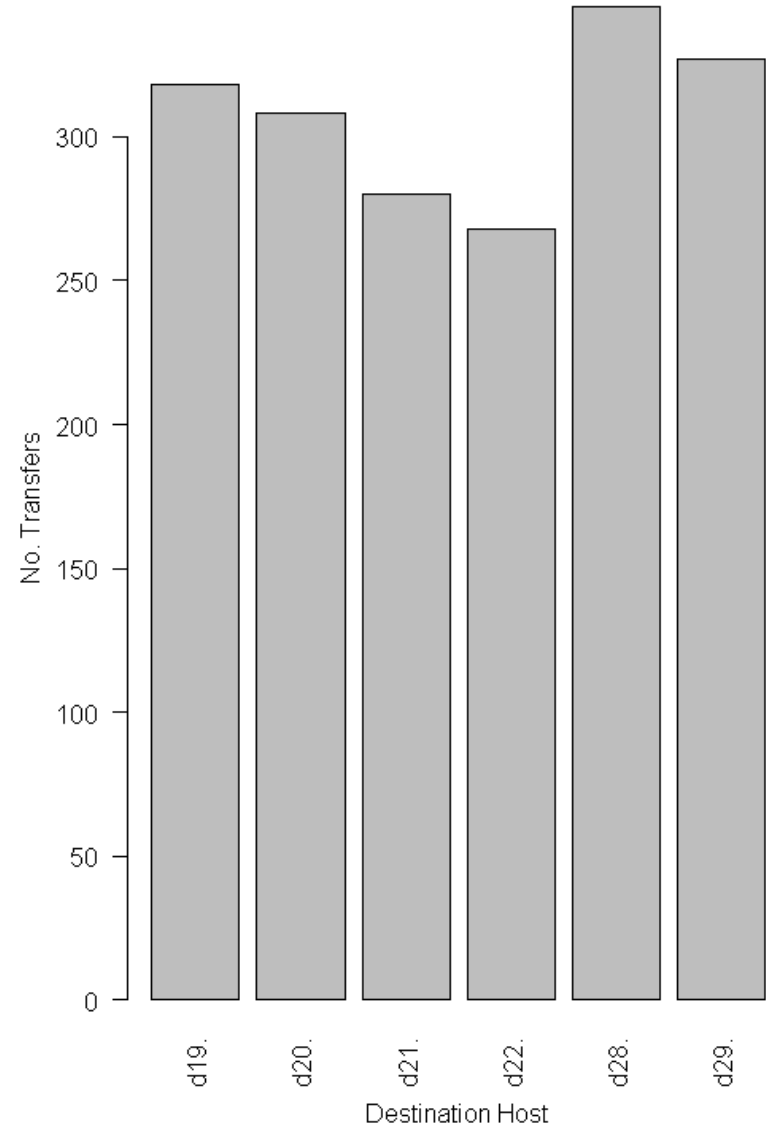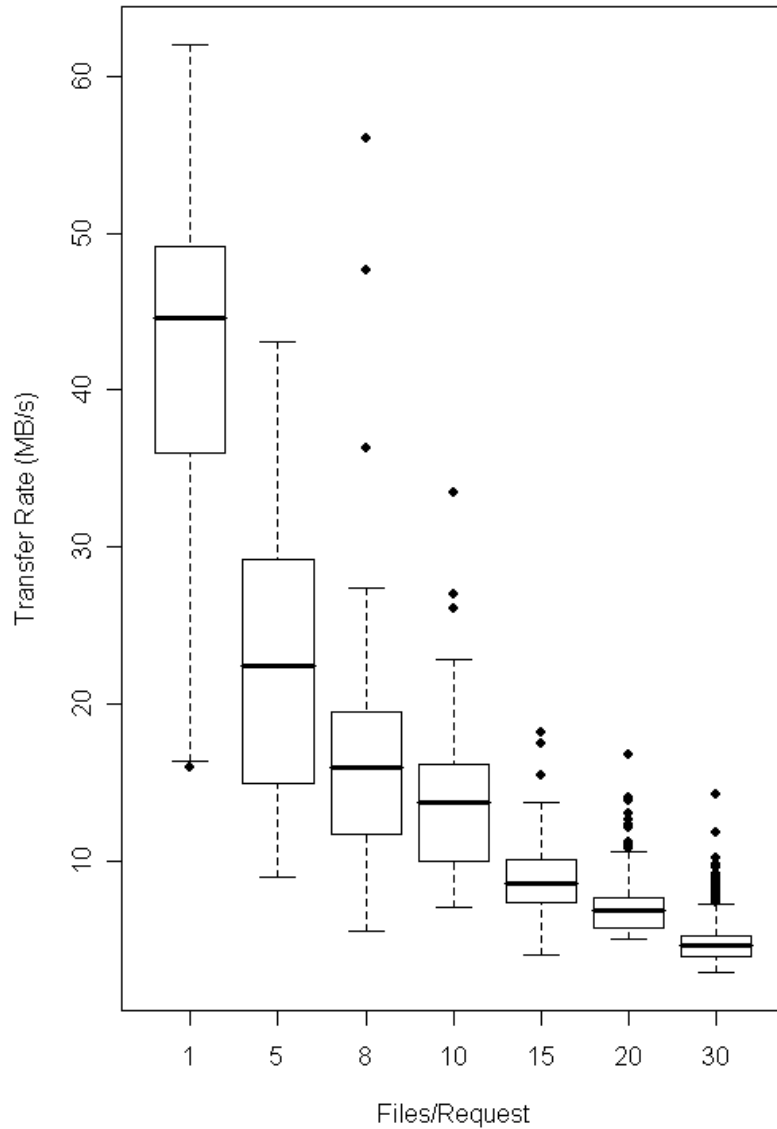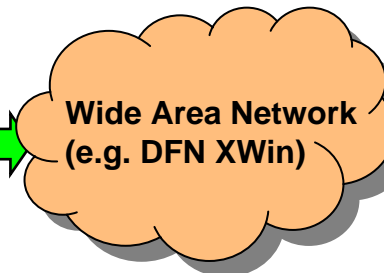

Load distribution over Pool Nodes

To be completed at DESY

- **Installation of ATLAS VO Box**
    - Scheduled for next week
- **Installation of latest LFC release and Migration of Catalog DB**
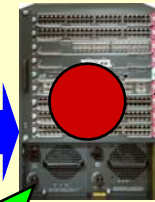- **Installation of ATLAS Software**
    - Scheduled for the end of this week

Shared Infrastructure provided by DESY Computer Center

DESY Infrastructure Services (Mail, Web, AFS, Backup, User Registry)
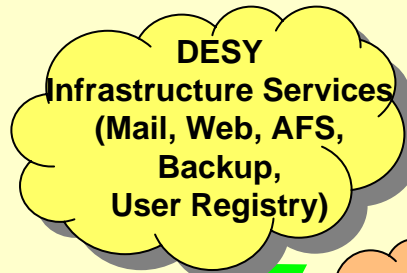
Wide Area Router and Firewall

Wide Area Network (e.g. DFN XWin)

General Computing Environment, Specific to the Tier-2 center

VO specific Infrastructure Servers - Calibr. DB - PhEDEx - …	Install Servers (Quattor)	Gateways/ Proxies
	Monitoring, Alarming *	Console Server

Tape Robots

DESY LAN

Proposed for 2007
• 10Gbps VPN => GridKa
• 1 Gbps shared

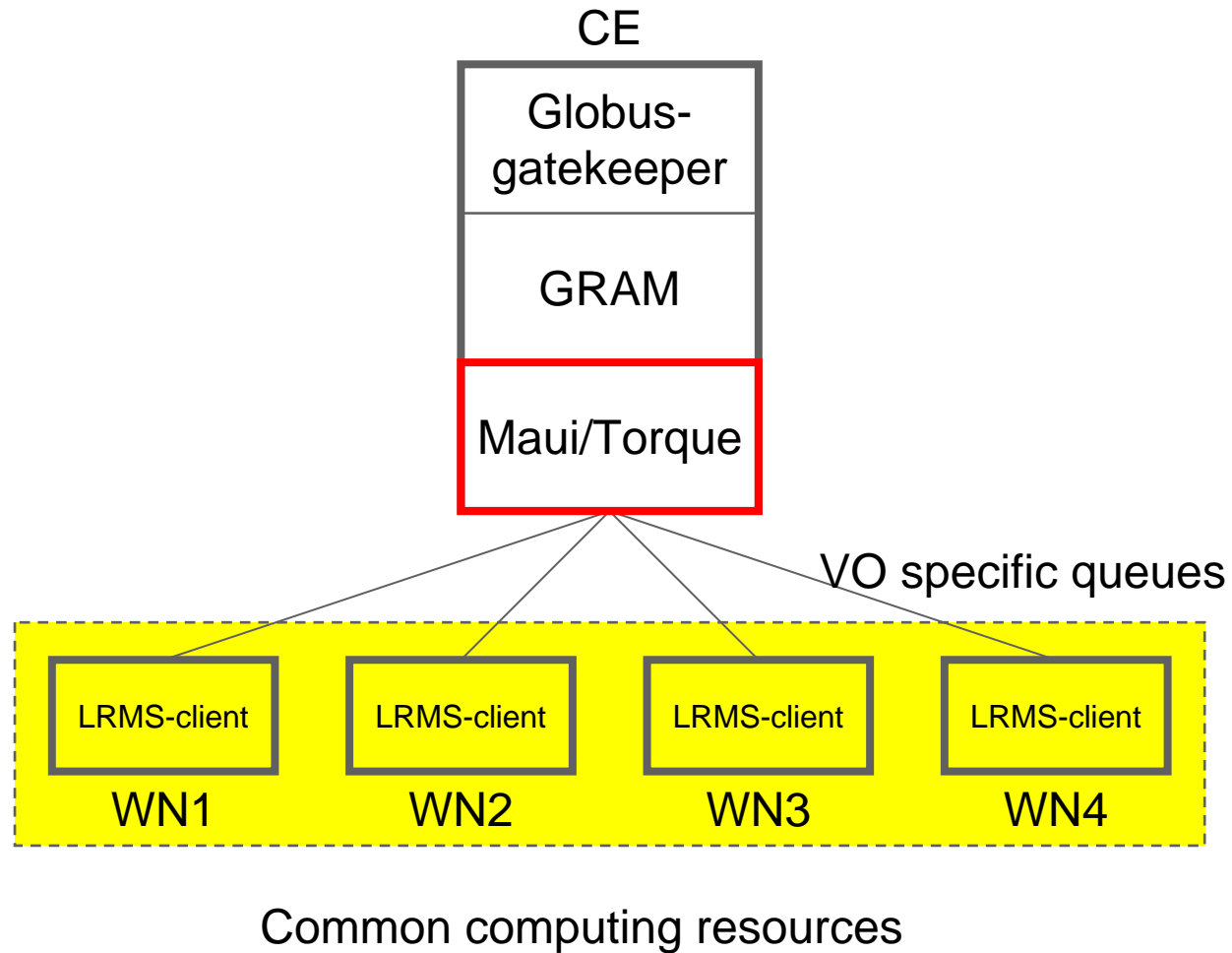*** Tier-2 sub-instance connected to central CC Monitoring/Alarming service**

Network Attached Data Servers
- **Autonomous Units built from Commodity Comp.**
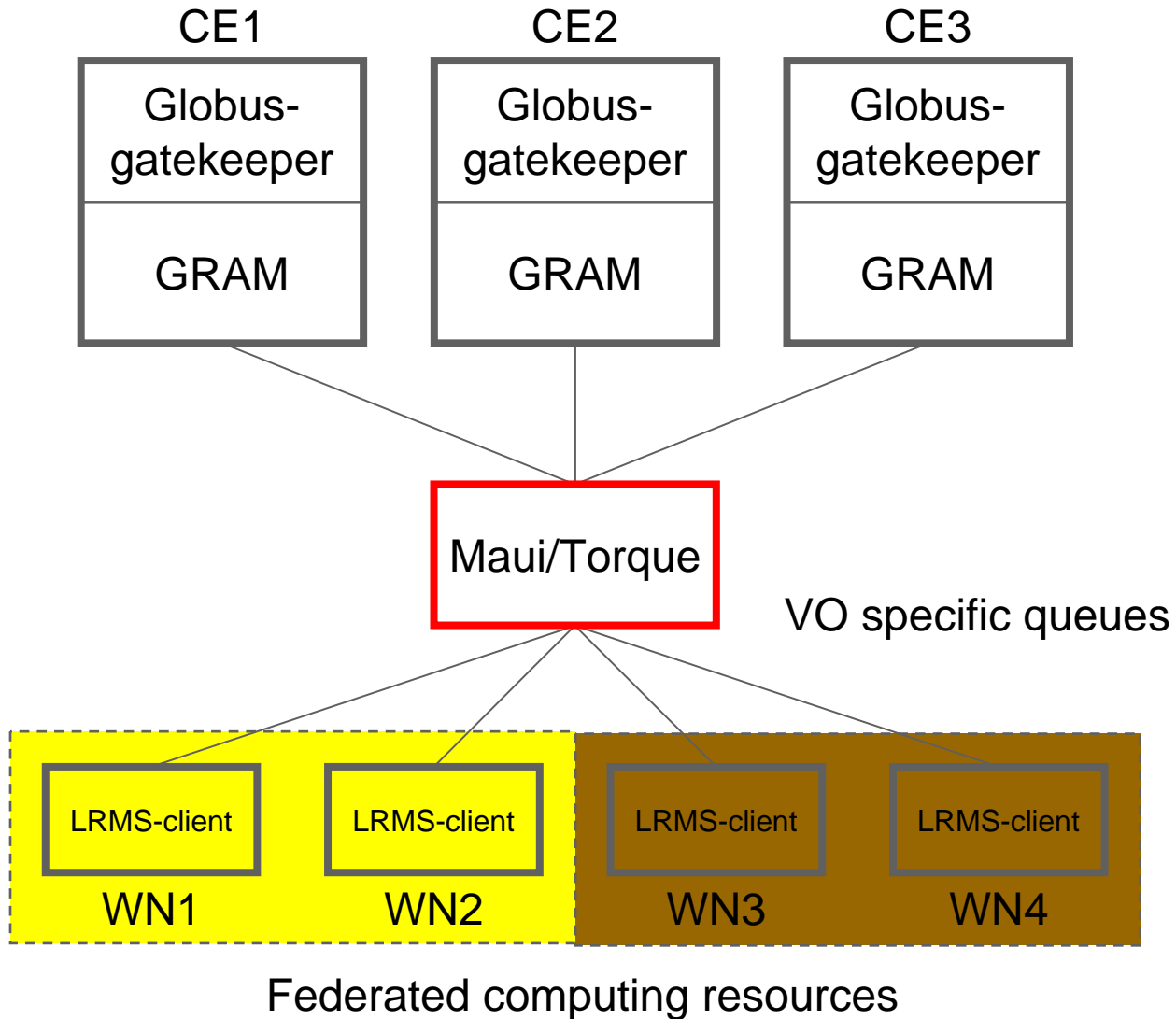- **SATA Drives ~ 10 TB each, Dual CPU Server, GigE connected**

Storage Managed by
- **dCache (Disk-only and HSM) for >90% of disk cap.**
- **Cluster File System for read/write acc.**

CPU Farm for
• **Selection**
• **MC Production**
• **Digitization**
• **…**

Network Switch > 300 GigE, 10GigE

Proposed LHC Tier-2 Center Architecture

Federating Resources …

CE

| Globus-gatekeeper |
| GRAM |
| Maui/Torque |

VO specific queues

LRMS-client   LRMS-client   LRMS-client   LRMS-client
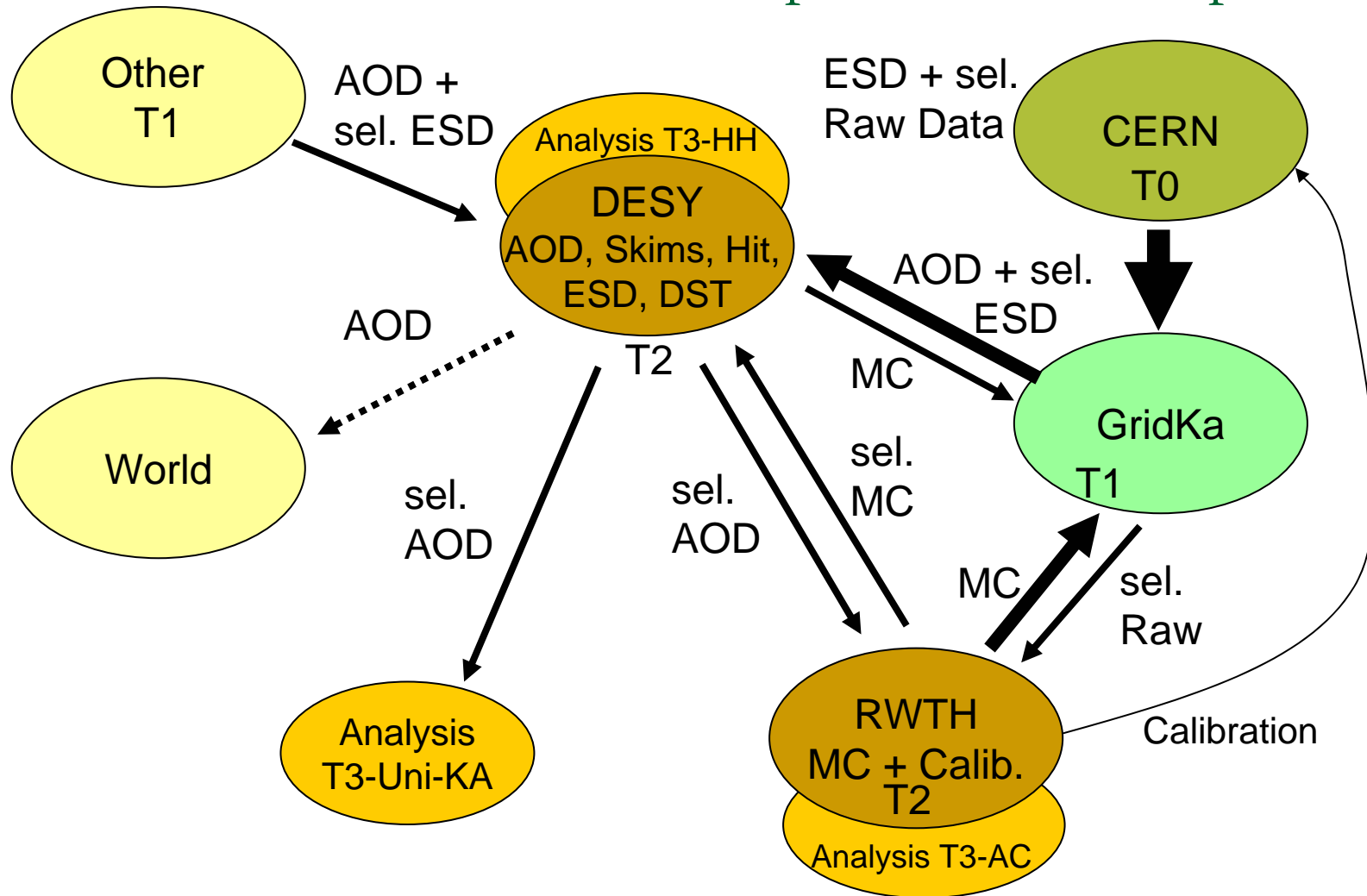
WN1   WN2   WN3   WN4

Common computing resources

… Federating Resources

A Federated Tier2 Center for CMS provided by RWTH Aachen and DESY – Operational Concept

Service profile at Sites

- **RWTH Aachen**
  - CMS Monte Carlo Production and Detector Calibration
    - Massive CPU
    - Moderate Disk Capacity
    - Tape Archive
    - Moderate WAN Bandwidth

- **DESY**
  - Data Processing (Selection), large scale Data Storage and Monte Carlo Production
    - Massive CPU
    - Massive Disk Capacity and I/O Bandwidth
    - Mass Storage Backend (Tape Silos) to Disk Cache
    - High Performance WAN Connectivity (10 Gbps Tier1 ⇔ Tier2 + 1 Gbps Tier2 ⇔ Tier3)
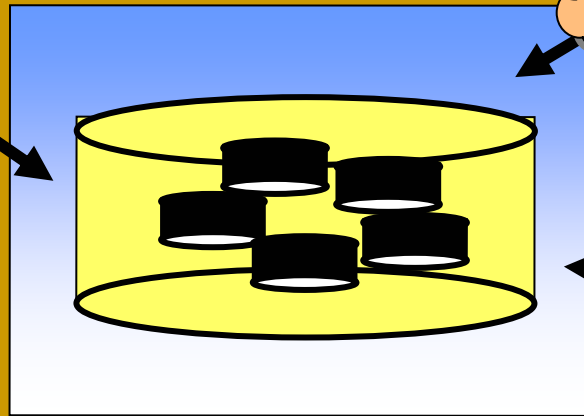
Grid Resources & Services at RWTH Aachen

- OS (common to all services): SL 3.05
- Quattor based install mechanism for all LCG nodes
- Quattor based install mechanism for dCache disk-only SE
  - In cooperation with DESY
  - Currently being tested

- LCG-2_6_0

- Site Services
  - GIIS
  - CE: Will be ~50 CPUs in Q1/2006
  - SE: dCache based, Classic SE (1,5 TB)
    - Storage: disk 15 TB in Q1/2006
  - UI: Automated RPM based installation on ~120 Institute owned Desktops
- VO-specific Services (CMS)
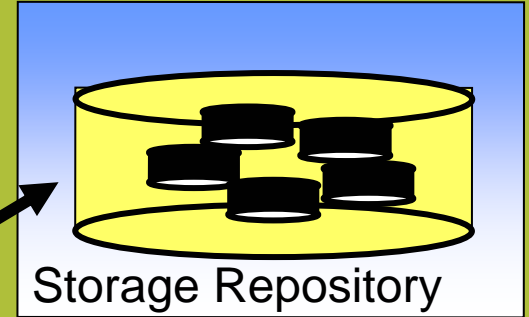 - PhEDEx (to be installed soon)

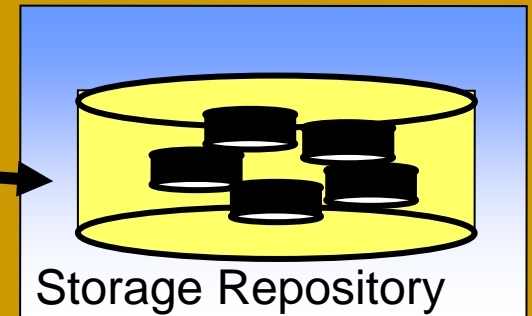# Option: Single dCache Instance w/ Storage Repositories in Multiple Locations

"Satellite" Cache at RWTH Aachen

Storage Repository

XWIN

Storage Repository

Central Cache w/ HSM support at DESY

Open Issues

- **Federated CMS Tier2 Center**
  - Job Submission via central CE or local ones?
  - Data and Storage resource sharing
    - File Catalogs
    - Single vs. independent Storage Systems
  - Sharing the expertise …
  - Managerial …
- **… many details to be worked out !**

Planned for SC4

- **ATLAS** DC related Data Transfers (FTS)
  - GridKa ⇔ DESY
  - DESY ⇔ other German ATLAS Tier 2 Centers
  - DESY ⇔ German ATLAS Tier 3 Centers
- Grid-based Job Submission & Execution using ATLAS Production System and new DDM for MC Production and Data Analysis
- Participation in Distributed DB Services (3D) for ATLAS

- **CMS**: PhEDEx based Data Transfers
  - RWTH Aachen ⇔ GridKa
  - DESY ⇔ GridKa
  - DESY ⇔ RWTH
- Grid-based CMS Data Analysis using combined Computing and Storage Resources at Tier2 and Tier3 Centers
- Grid-based MC Production (MCPS)
- Planning to share manpower resources
  - RWTH will send person to DESY to help w/ SC4 in Spring

Expectations from Tier1

- **Immediate access to new and archived data**
  - i.e. AODs, selected ESDs and Raw Data
- **Access to MC Data stored at Tier1**
- **Reliable and low latency pass-thru of data transferred via Tier1 (e.g. according to PhEDEx routing policy)**
- **Reliable Network Connectivity w/ sufficient bandwidth**
- **Clear, low latency communication channel (incl. escalation procedure and tracking)**