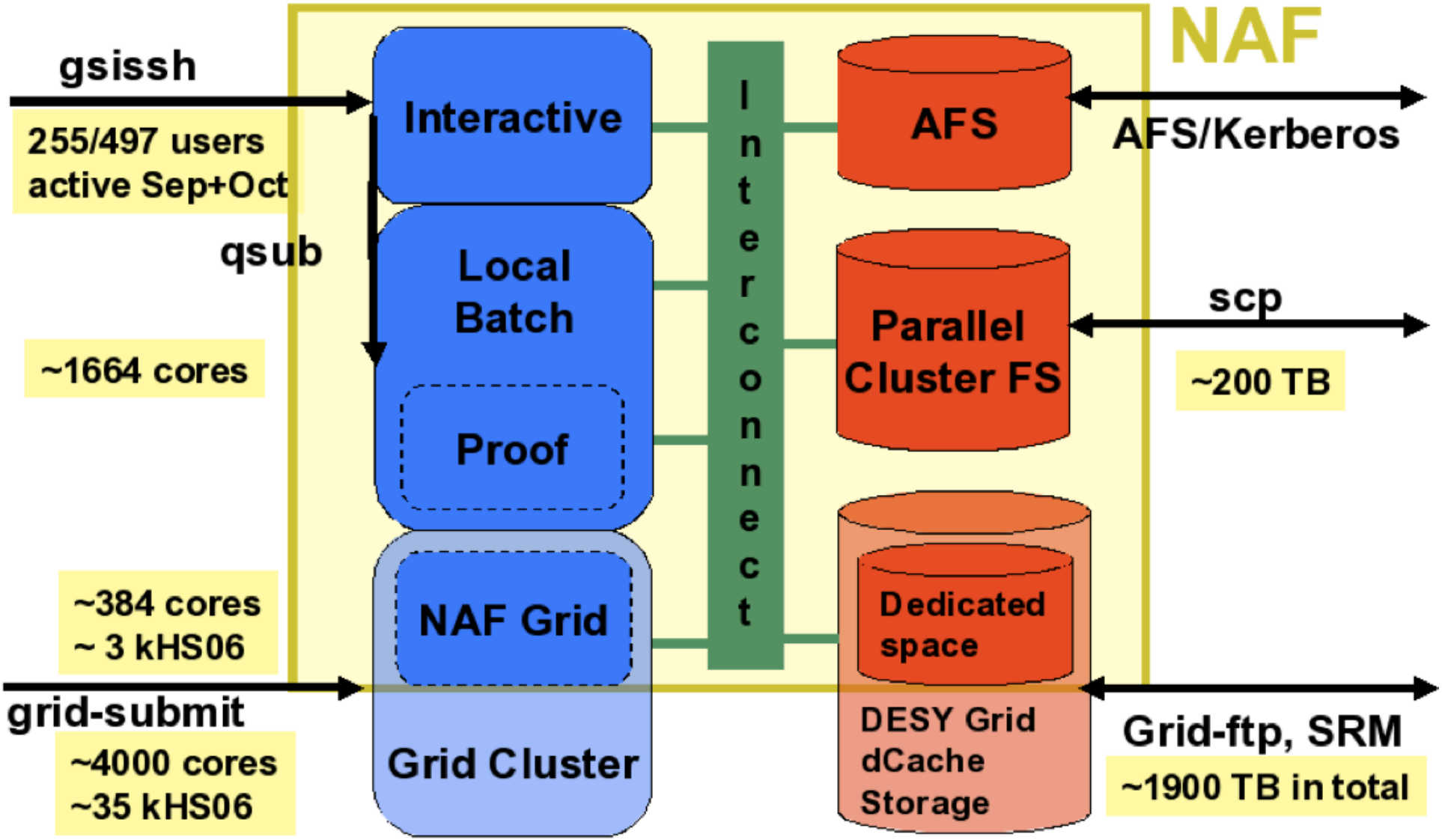
NAF & NUC reports

Y. Kemp for NAF admin team
H. Stadie for NUC
4th annual Alliance Workshop
Dresden, 2.12.2010

HELMHOLTZ
ASSOCIATION

DESY

NAF: The Overview Picture and current resources

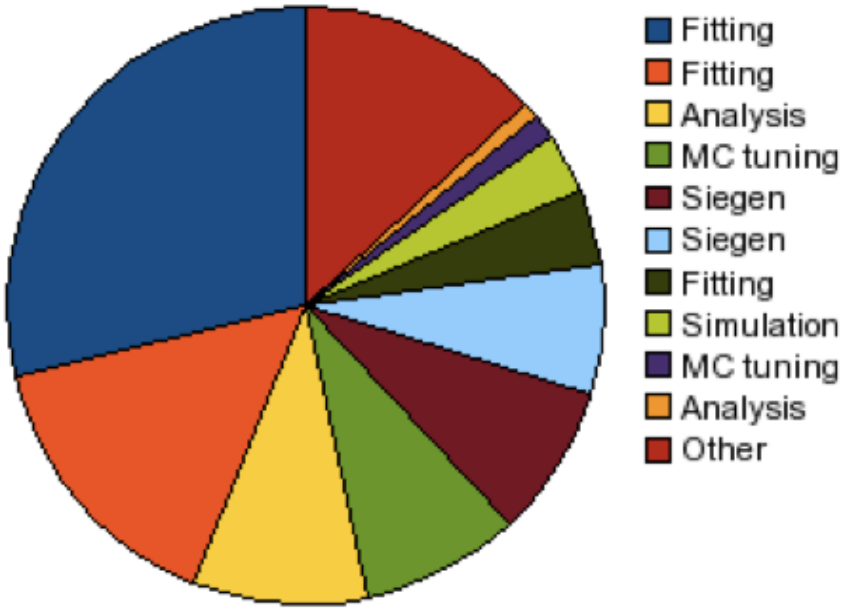


Hardware is distributed over two locations: HH and Zeuthen

ATLAS NAF CPU usage

- Current snapshot of ATLAS NAF batch usage
  - accounting interval: 1.8. - 29.11.2010
  - total CPU time: 21362 days
  - ➔ ~12% of NAF CPUs
  - total wall clock time: 58228 days
  - ➔~33% of NAF CPUs
  - nominal ATLAS share: ~25%
  - ATLAS is working on the NAF

CPU time fractions
NAF ATLAS



- Analysis type jobs are becoming more prominent within the group of power users
  - Other category includes ~100 users
  - As an example: September 2010 - 71 users total, 28 from DESY/HUB

(slide provided by M. Barisonzi & W. Ehrenfeld)

ATLAS NAF disk usage

- The ATLAS dCache storage consists of
  - T2 pledge storage 2010: 740 TB total, 489 TB used (66%)
  - NAF/DESY storage: 441 TB, 303 TB used (69%) --> only 1⁄2 T2
- Current lower usage due to
  - on going data reprocessing
  - on going MC processing
  - some dedicated space given to ATLAS groups: 150 TB total, 104 TB used (72%)
  - new pledged space will not be available before 1st April 2011
- NAF/DESY storage is used for:
  - extending the DATADISK/MCDISK space tokens to have the full AOD set available 260 TB (need more space in DATADISK)
  - provide user permanent space in LOCALGROUPDISK space token: 175 TB used for

    User data
    Group data not officially available at DESY T2 (group ntuples)
    archive of older data/MC if needed
    High demand, current usage is 82%, more space needed

NAF usage by CMS

- CMS:
  - Install CMSSW on NAF AFS
  - Adapt submission frameworks to local batch
  - Jobs access data on Tier-2 dCache SE
  - Interactive data analysis with PROOF and Lustre
- CMS: Additional data sets (160 TB) at DESY
  - All data very well used by community, often many users per dataset
- Tasks performed:
  - (Prompt) data analysis
  - Special MC sample production
  - Development of analysis tools
  - Calibration, alignement,…

CMS Physica Analysis Summary:

CMS PAS BPH-10-002

CMS PAS JME-10-004

CMS PAS QCD-10-005

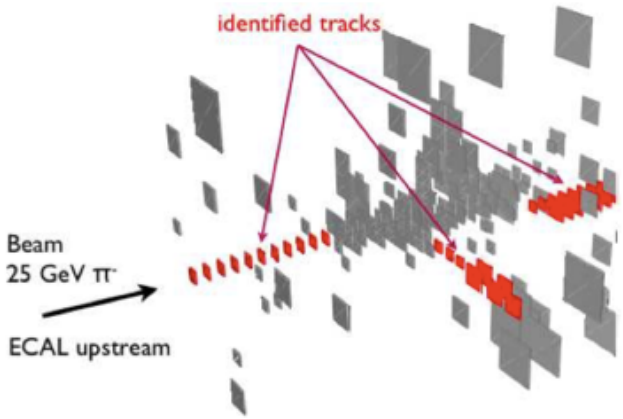
Acknowledgements

We would like to thank Matteo Cacciari, Gavin Salam and Sebastian Sapeta for their precious help in understanding the theory and their useful suggestions. We also thank the National Analysis Facility (NAF) administrators and WLCG for providing the excellent and reliable computing infrastructure necessary to carry out this analysis.

Extract from CHEP 2010 presentation by Kai Leffhalm

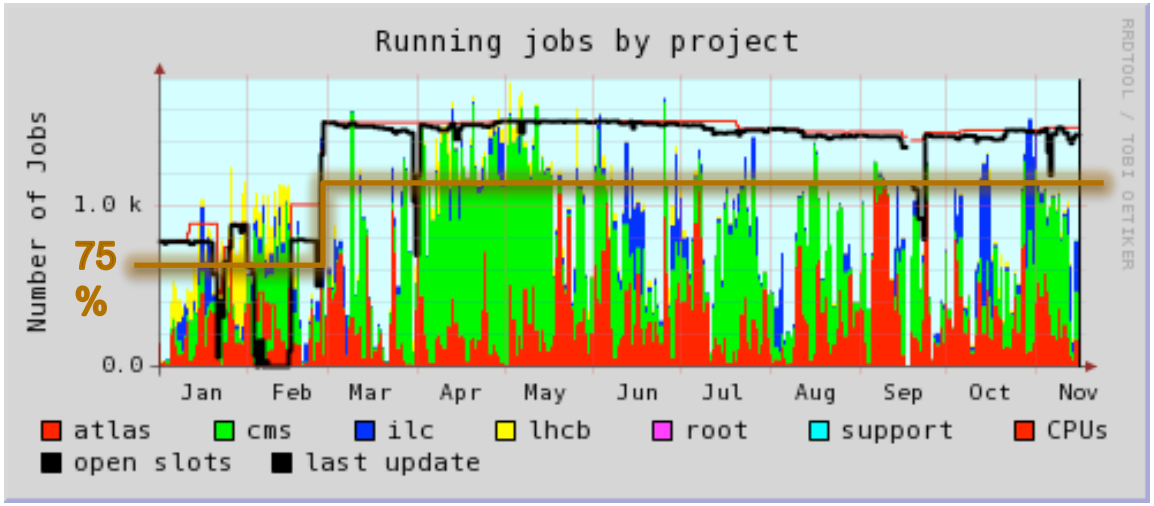
NAF usage by LHCb & ILC/CALICE

- LHCb:
  - E.g: Study of CP-violation in the B sector: Requires complex max. likelihood fits
  - Generate “toy” MC, very CPU intensive, fast turnaround, short jobs
  - most users perform nTuple production
  - LHCb uses resources as expected
  - NAF important pillar of their analysis infrastructure
- ILC:
  - ILD LoI: Studies of impact of machine background on track reco efficiency
  - Fast turn-around time for efficient prototyping
  - NAF: Easy to manage jobs
- CALICE:
  - GEANT4 validation with AHCAL data
  - Custom MC generation
  - NAF: work with scripts in homogeneous environment
  - … and keep efficient access to Grid storage



Extract from CHEP 2010 presentation by Kai Leffhalm

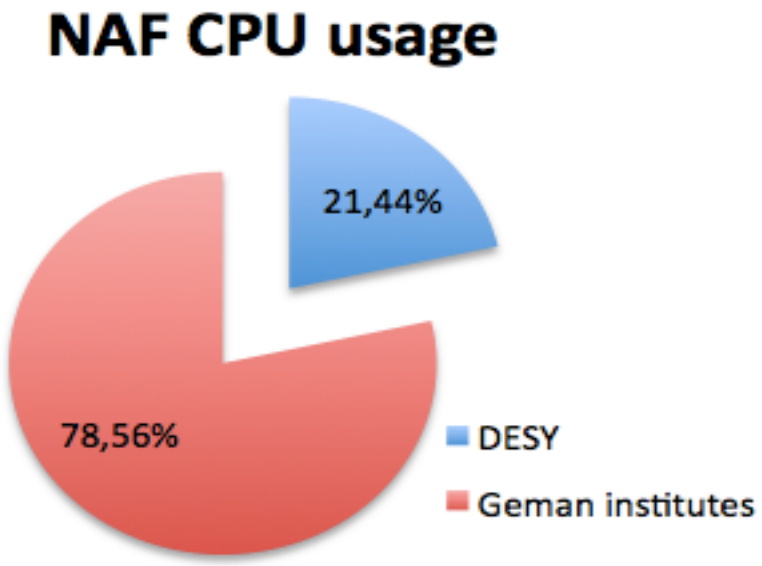
NAF Resources well used



Need upgrade in 2011!

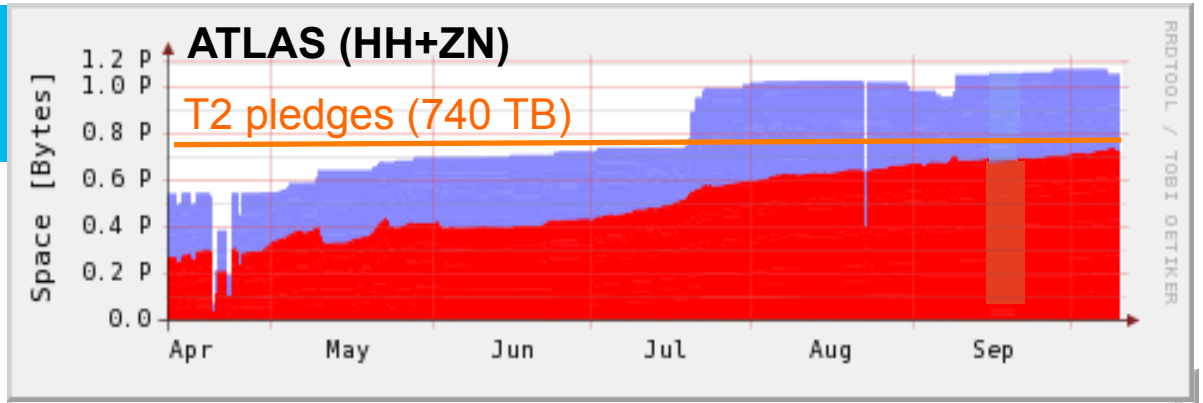
Recommended limit: < 75%, peaks up to 90%

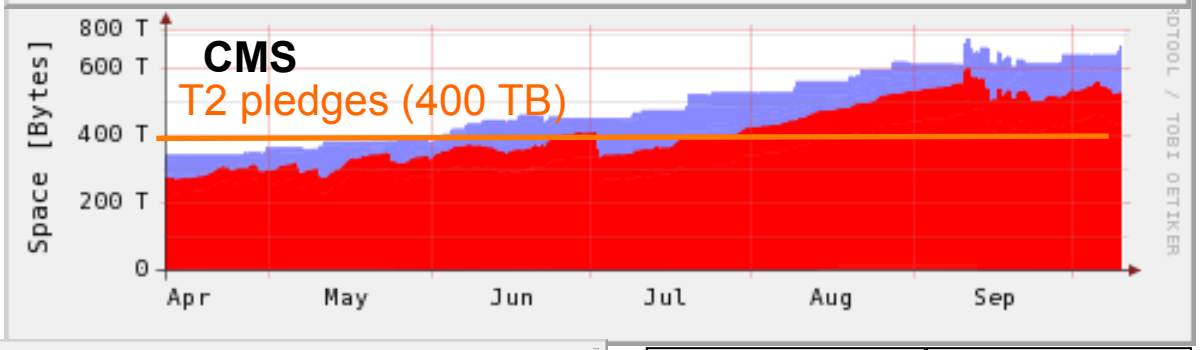
- NAF well used by German institutes
  - 21% used by DESY scientists.



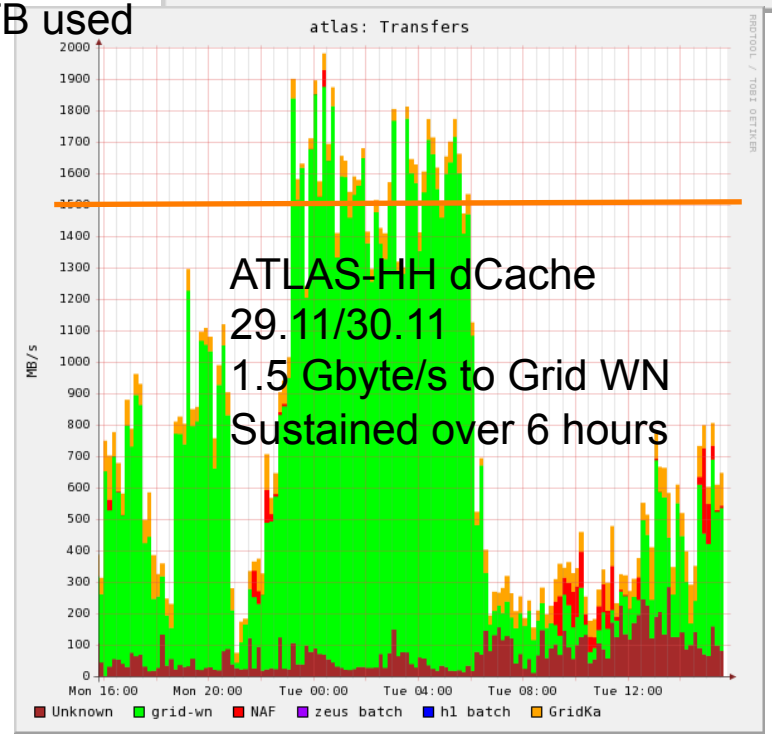
dCache storage & NAF

- Both ATLAS and CMS have substantially more space in dCache compared to T2 MoU pledges
- NAF and user space
  - And other contributions, e.g. UniHH-CMS
  - E.g ATLAS: T2 part 66% used, NAF part 303 TB/441 TB used





Used space | Free space



After observing data taking for ~one year now:

1) Optimize dCache for speed
2) Optimize dCache for safety and availability of “custodial” data
3) Optimize dCache usage and data placement for non-T2 data

dCache is *THE* working horse for data storage

Hardware Status

NAF is three years old now

- Have to start replacing first hardware
  - First replacement currently ongoing
  - Newer hardware, more RAM/core, new network technology,…
  - At the end, more computing power
  - … 10 Gbit infrastructure
  - … and more to come in 2011
- New additions to dCache storage (quantity & quality)
- **Clear commitment from DESY to support NAF**

- Future purchases planned together with the NUC and take into account findings of the GridCenter Review Task Force.

Problems and Issues

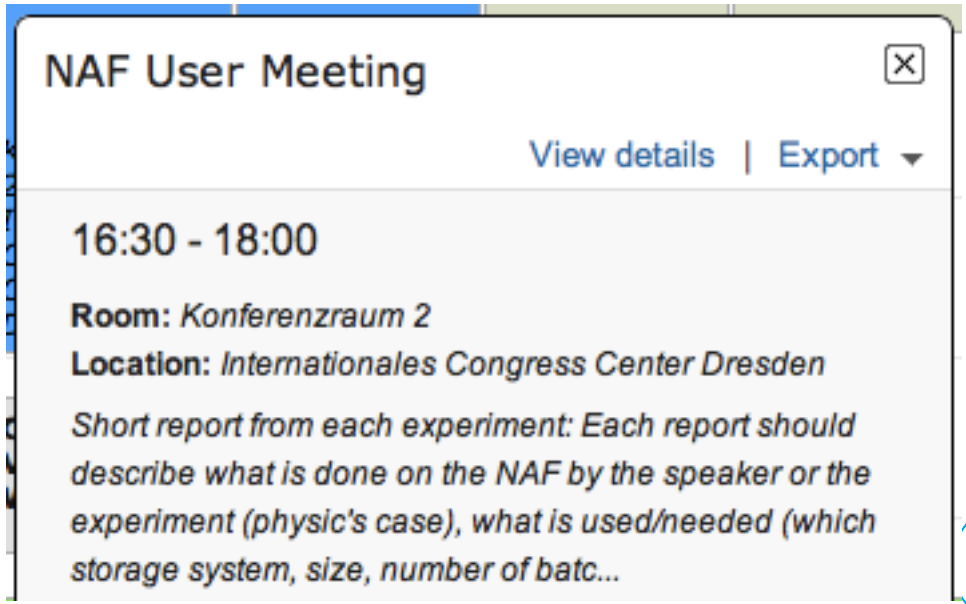
- Problems started ~Mid July: Whole AFS instance unavailable for some minutes at a time
- Debugging difficult: Consulting with AFS developers
- Main cause: SGE behavior with NAF job type when starting many jobs at the same time
- First countermeasures taken, more to come
- User training will start this afternoon

- Many features still not working reliably (e.g. group quotas, ACL,…)
- Maintenance tools to make users' life easier not yet available (deletion tools,…)
- Overall stability improved, but some hick-ups are still seen
- Performance reports unclear – no end-to-end performance investigation done
- Future of Lustre unclear – in general (ORACLE) and at DESY: looking for alternatives
- The need for such an "easy-access" large file store is indisputable

DESY

NAF User Committee – and User Meeting

- Monthly meetings of the NAF User Committee. Members:
  - ATLAS: Marcello Barisonzi & Wolfgang Ehrenfeld
  - CMS: Andreas Nowack & Hartmut Stadie (Chair)
  - LHC-B: Johan Blouw & Alexey Zhelezov
  - ILC: Steve Aplin & Shaojun Lu
  - IT: Andreas Gellrich & Kai Leffhalm
- … status reports and discussions with NAF technical coordinators
- NAF Users Meeting
  - … see you there!



“Random comments” collected by NUC

- “the currently available resources, especially CPU in the batch system, could provide good working conditions, when all systems are working properly”
- “ongoing problems make an effective and timely data analysis almost impossible”
- “dCache user diretories are not reliable enough”
- “congested work group servers”
- “slow I/O with dCache” (data placement), “need more space”
- “add more Lustre space”

NUC: Some words on support

- Support Ansatz: Two different paths
  - Problems with central NAF services ➔ DESY helpdesk
  - problem with experiment infrastructure ➔ experiment mailing list
  - (+ second level support structure, available for experiment experts directly)

- Challenges:
  - dedicated manpower for central services?
  - dedicated manpower for experiment support? (FSPs)
  - O(Min) response time?
  - analysis with fast turn-around needs very reliable system (better than Tier-2 MoUs)