NAF Status

- Where are we?

Yves Kemp, DESY IT

8th Terascale Alliance Workshop
DESY 1.12.2014

- This is an interactive presentation: You can steer the direction
- This presentation contains, in this order:
  - Problem reports for the expert and advanced user
  - Some personal thoughts on status and future
  - Some questions to users
  - The slides from last year: For those who want to know what the NAF is, and what NAF 2.0 is (and NAF 1.0 was)

- Vote: What should I present?
  - … one alternative being not to present anything and just do a Q&A round

- … problem reports

NAF 1.0 decommissioning

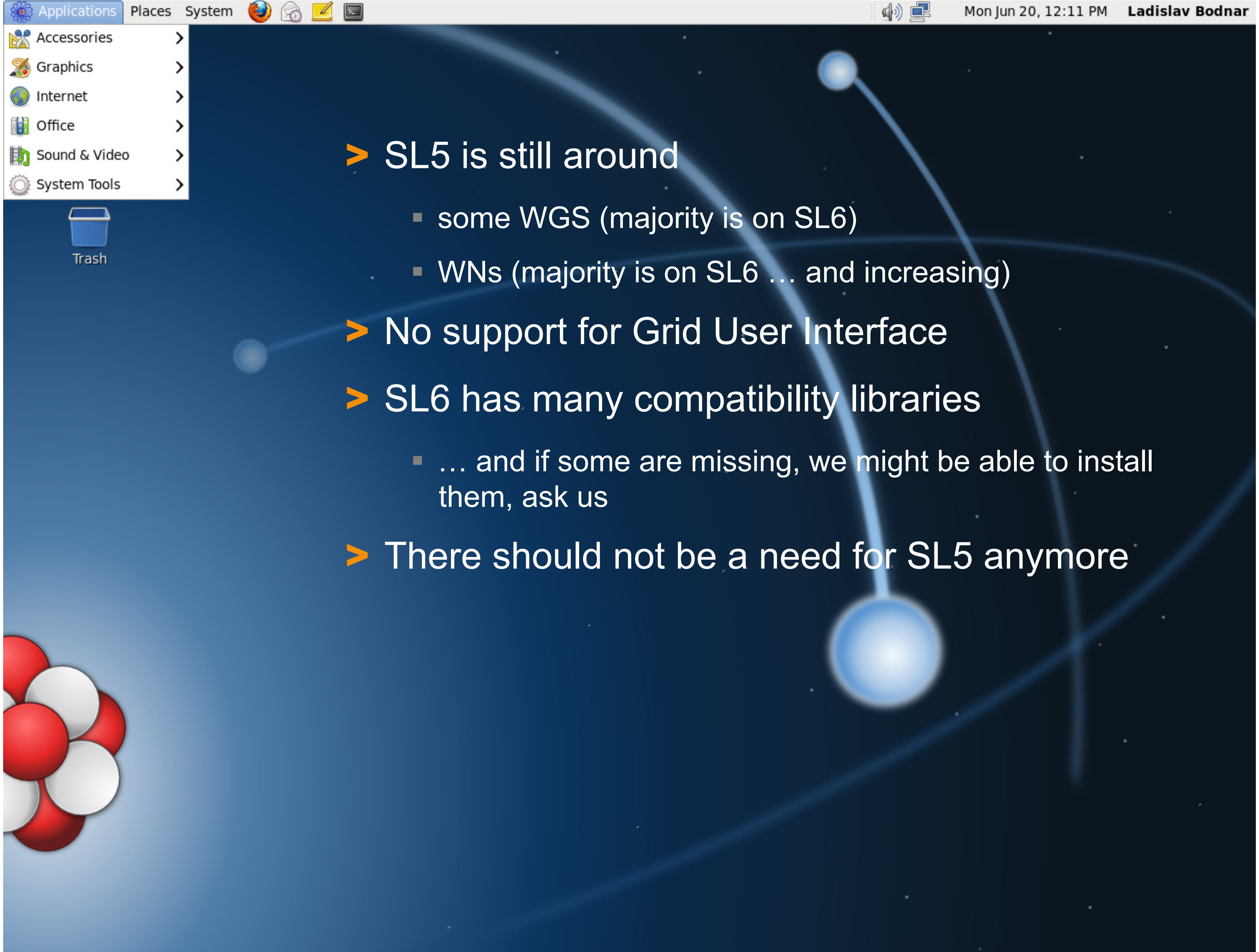
- Most batch migration done end 2013
- NAF 1.0 decomissioning April 2014
- SONAS fileset migration mostly done early 2014
- NAF 1.0 AFS decommissioned mid 2014

- Decommissioning basically worked without problems

The CVMFS story

- Started 2014 with CVMFS-over-NFS mounts for all experiments
  - Worked basically OK
  - … except that ATLAS reported performance issues
- Tested one client with native CVMFS
  - Performance was much better
- Investigated on how to perform larger deployment
- Now: Whole DESY-HH Grid has migrated from CVMFS-o-NFS to native CVMFS
- All experiments WGS are on native CVMFS
- New BIRD/NAF WNs will be installed with native CVMFS
- … which leaves SL5 machines on CVMFS-over-NFS
  - Caveat: CVMFS-over-NFS now is deprecated at DESY, and running on unsupported hardware
  - Until now, no major problem observed with native CVMFS

SL5 issues



- SL5 is still around
  - some WGS (majority is on SL6)
  - WNs (majority is on SL6 … and increasing)
- No support for Grid User Interface
- SL6 has many compatibility libraries
  - … and if some are missing, we might be able to install them, ask us
- There should not be a need for SL5 anymore

And SL6?

Scientific Linux

- We currently have two SL6 flavors: Salad configuration and Puppet configuration
- Some things differ in the two configuration methods
  - RPM: … if you are missing an RPM, contact Helpdesk
  - Login scripts handling
  - Setting of environments
- In Puppet, we opted for an environment as close as possible to the original distribution
- … you might need to explicitly source settings
- RedHat Enterprise Linux 7
  - CentOS 7 vs ScientificLinux 7 no technical difference – will decide after 7.0->7.1 update

Supporting schools and workshops

- Remember: NAF 1.0: You needed a certificate to log in.
- Usually the most difficult thing to organize for schools

- NAF 2.0: Normal account+password
  - Easy: DESY has school accounts
- Futher infrastructure:
  - Easy to deploy handful WGS dedicated to school (have done for up to 6)
  - Configure special share and priority in batch if needed
  - Temporary space on AFS or SONAS
- Take-home-message: Ask us if you plan a school in the Terascale context and need a computing infrastructure

Batch

- BIRD (the batch facility for NAF 2.0) currently is the largest batch installation at DESY in numbers of WNs
- Setup rather complicated: Many different user groups and usage profiles, …
- Have used Sun Grid Engine (SGE) in the past, as this is discontinued, using Son of Grid Engine (SoGE) … but we can still call it SGE
- SGE has shown scalability and stability issued. Performance issues are either gone or users used to them…

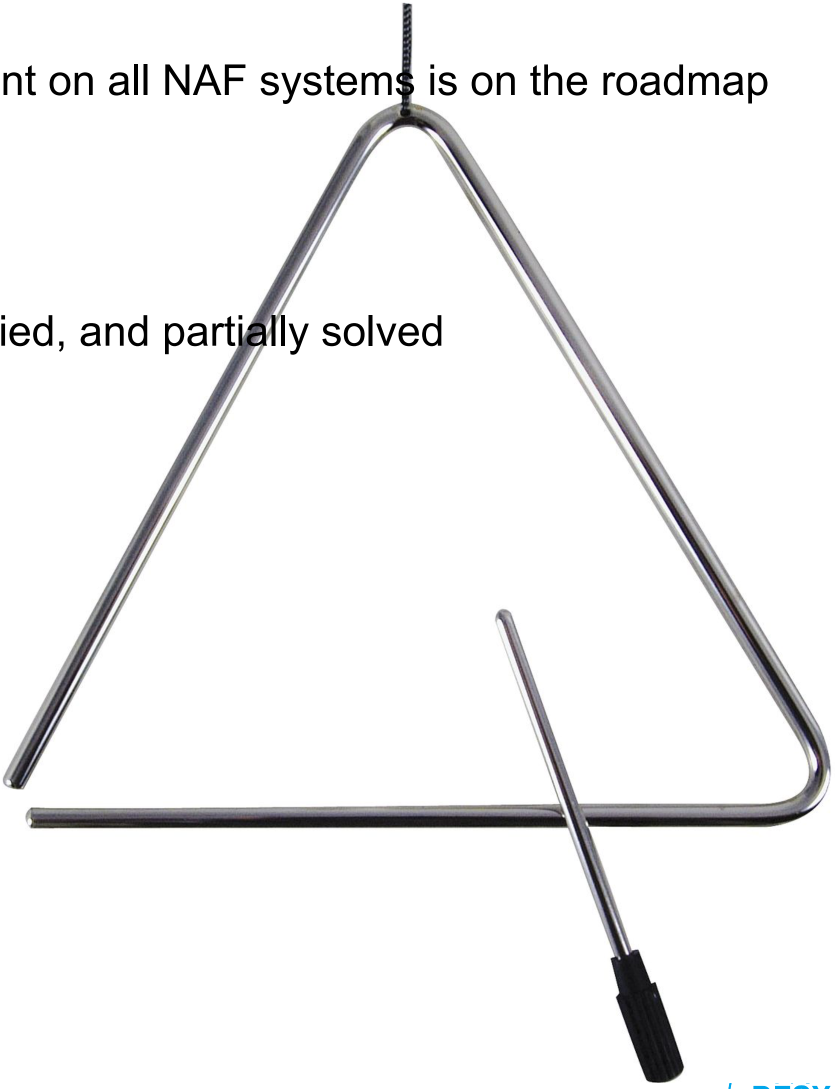
- We plan for renovating batch system in 2015
  - Must-haves: AFS&Kerberos support, fairshare and priorities, single&multi-core jobs, parallel jobs, eventually something like array jobs,…
  - Will differ from current setup in some points … will present a “translation table”

Monitoring & Documentation

- Batch: http://bird.desy.de/status/
  - Further work on batch monitoring once “the new batch setup” is clear
- WGS:
  - Long-running jobs: Users are alerted as well as group responsibles
- General: Still planning to deploy Ganglia, awaiting results of internal monitoring task force
  - Ganglia is good, but will not cover all metrices needed by IT for internal monitoring
  - Plan to make (some) plots available to NAF users
- NAF: Now has a wiki: http://naf-wiki.desy.de/
  - Users are invited to add what they think is missing

Storage?

- dCache: No major issues
  - Maybe one thing to report: NFS 4.1 mount on all NAF systems is on the roadmap
- SONAS:
  - Mostly migration related issues
  - Observed one major bug in 2014, identified, and partially solved
- AFS:
  - No major issues reported in 2014
  - XXL volumes major AFS topic



- … some personal thoughts on status and future
- … some questions to users

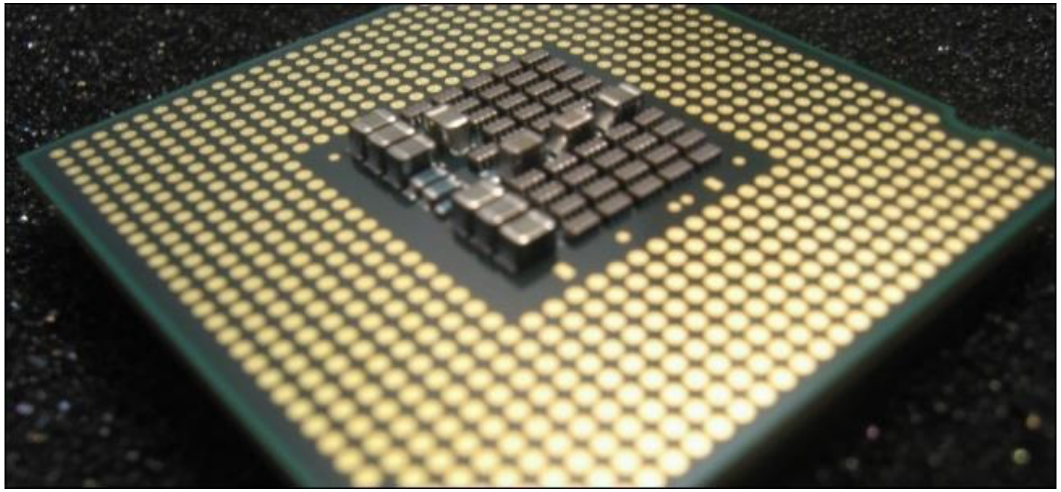
- NAF 1.0 -> 2.0 transition was necessary
- NAF 1.0 -> 2.0 transition was not always easy
- NAF 1.0 -> 2.0 transition was successful in the end

- NAF 2.0 is working and fulfilling its role in the German LHC landscape

- What do we/you need for NAF 2.1 ? Or NAF 3.0 ?

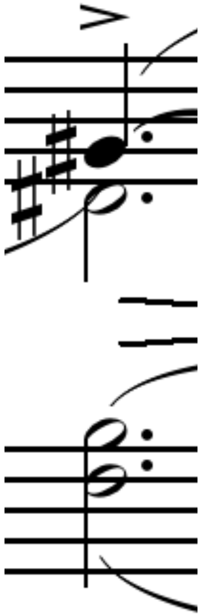
CPU: WGS + WN

- Currently, well setup
- LHC analysis currently well served with (few) interactive WGS and (lots of) batch WNs
- LHC analysis currently well working with defined operating system
  - SL6 currently and probably also in 2015
  - Do not see "cloud"-like CPU needs for analysis in NAF context for foreseeable future



- "NAF remote desktop"
  - Little use, but has word spread?
  - FastX now is the product behind this. Situation has cleared – and this is a magnificent tool!

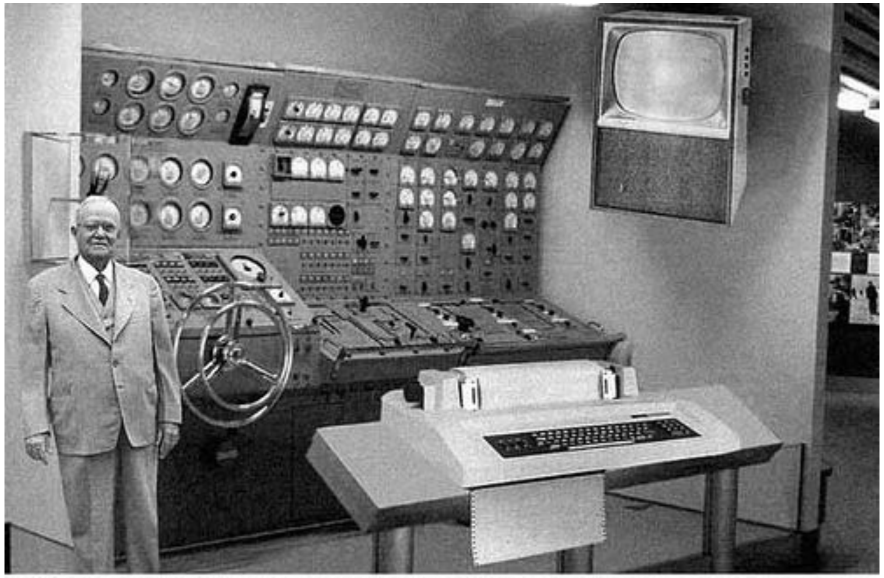
Storage:



- Storage triad:
  - dCache for the mass data
  - SONAS for the fast data
  - AFS for the small and world-wide accessible data
  - (+AFS-XXL for the not so small data)
- … will remain with us for some time
  - Although not always easy to communicate and work with
- Evolutions on the horizon (my view):
  - dCache: More protocols (NFS v4.1, WebDAV, Cloud-Access …)
  - SONAS: “Just an NFS server” … will change product in some future … nothing to worry in my opinion.
  - SONAS (or successor) long-term-perspective: Superseeded by dCache? Replace AFS?
  - AFS: Has shown some deficiencies. No easy replacement now, but will see again in ~five years
  - ... What do you need? What don’t you need?

New stuff

- DESY has a HPC cluster
  - Some kind of proposal workflow: Write us an email and we will check whether we give you access
  - Idea: Plattform for massive parallel jobs with large communication needs
  - .. That explicitly means: NO PROOF!
  - I could imagine some HEP workflows are there that can benefit from such a plattform
- DESY has some GPGPU machines
  - Access is similar as for HPC cluster
- New developments in [G|C]PU
  - ARM, Intel MIC, Power8, … anyone?



Scientists from the RAND Corporation have created this model to illustrate how a "home computer" could look like in the year 2004. However the needed technology will not be economically feasible for the average home. Also the scientists readily admit that the computer will require not yet invented technology to actually work, but 50 years from now scientific progress is expected to solve these problems. With teletype interface and the Fortran language, the computer will be easy to use.

- … the slides from last year

NAF 1.0: How did it go?

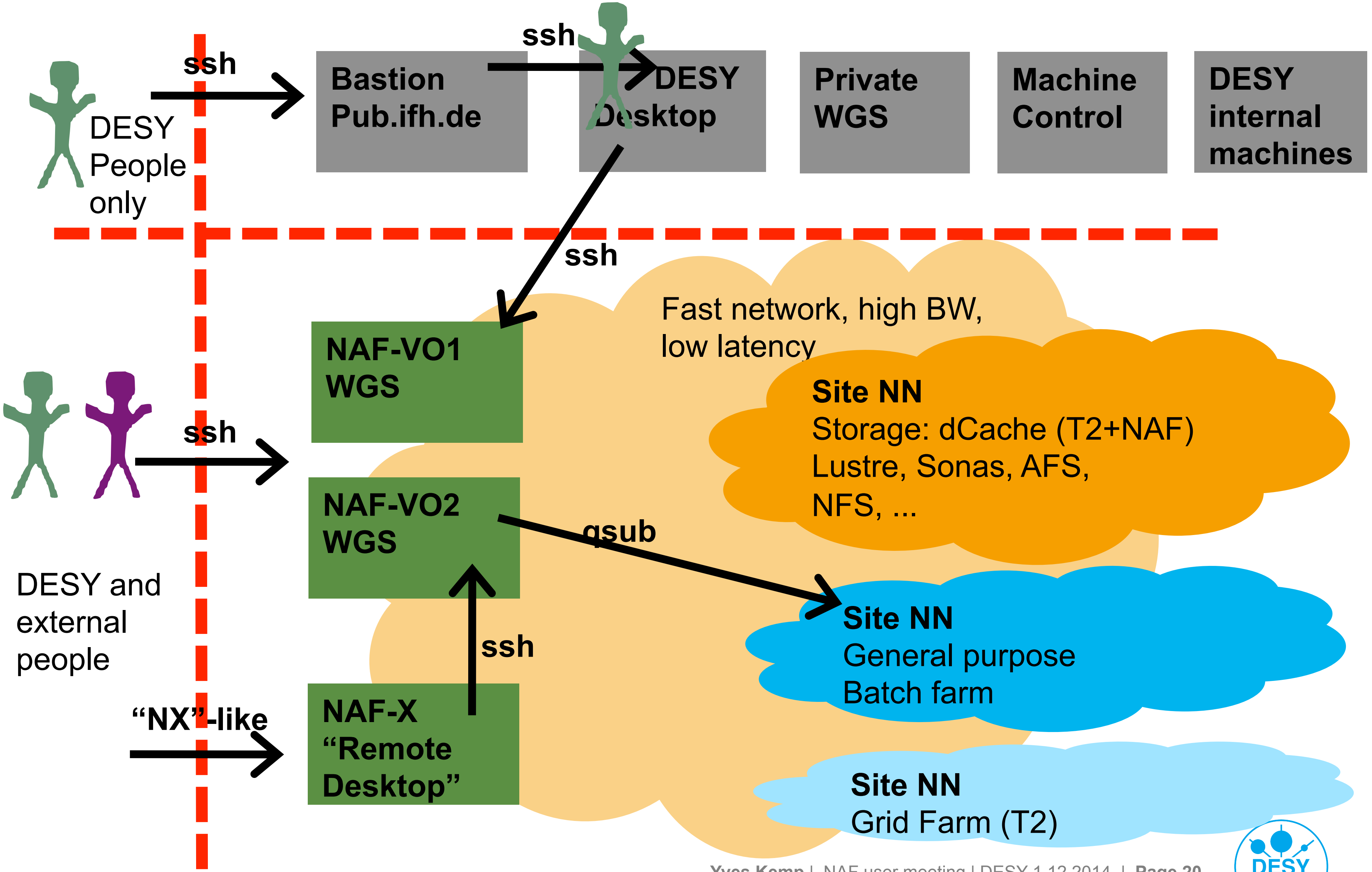
- Fast setup, many users, many successful analyses. General setup OK.
- Identified some weak points though:
  - Missing integration into “normal” DESY proved to be manpower intensive and decoupled NAF from advances in “normal” DESY infrastructure
  - A further spread to other German HEP sites did not happen
  - We never benefited from the two-site setup – actually, we suffered from it in terms of reliability and performance (latency!)

- Needs have evolved since 2007
  - More graphical tools needed
  - More software, e.g. also commercial one
  - General mobility ask for better remote capabilities
- **The NAF needs a fundamental redesign to continue its success story**

Future of the NAF: NAF 2.0 - What?

- Everyone will get a “normal” DESY account
- Will have access to a restricted set of “normal” DESY resources
  - The NAF 2.0 resources
  - Including several WGS
  - Including large batch system
  - Including $LargeFileStore (e.g. Sonas)
  - Including dCache access
  - …
- Technical details:
  - Closer integration into respective site (HH or ZN)
  - No data should go over the WAN - This is enforced in case of $LargeFileStore
  - Plain ssh+Password login – gsissh planned for later
- Support: Better integrated into site support, more people know infrastructure
- New developments
  - … also new communities (belle, herafitter, …)

Relation NAF 2.0 <-> DESY infrastructure



NAF 2.0 and the two DESY sites

Standort Hamburg »
Anfahrtsbeschreibung

Standort Zeuthen »
Anfahrtsbeschreibung

> At the heart of every analysis is the data

- fast and reliable access is the most important design criterion for the NAF

> From the four NAF experiments ATLAS, CMS, ILC and LHCb, three have storage concentrated at one DESY site: CMS, ILC and LHCb

- Having CPU resources at the other site makes no sense

> For ATLAS, the DESY directorate has decided to provide enough resources for NAF usage at the Hamburg site

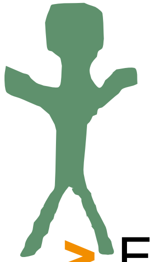
- ATLAS will only use the Hamburg site for NAF 2.0

> For CMS (as well as ATLAS and ILC), **NAF 2.0 will be Hamburg-only**

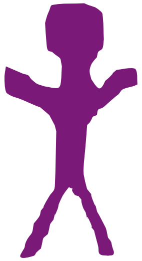
- **Profit from the close integration with Tier-2 Storage!**

Status: Getting an account

- Prerequisite for getting an account: Being known at DESY as a person
- We need to enter you in PIP system: PersonenInformationsPool
  - Name, Firstname, Affilitation, Date&Place of birth, … and this needs to be accurate, unique and somewhat certified
- DESY people (or Uni-HH, Humboldt Uni and other befriended institutes)



  - You are already registered in PIP !
  - Your normal DESY account (the one you use for bastion.desy.de) just needs the resource “batch” in the registry – and there you go
- External people



  - You need to get registered in PIP
  - Web-based registration form set up – certification by your Grid certificate
  - The PIP entry and the account creation is then done “automagically”

When / how should users migrate?

- NAF 2.0 is ready
- You can register now – and start working

- NAF 1.0 will continue to work for some time – but resources reductions already has started!
- After 1.1.2014 major work should happen on NAF 2.0
- We want to shut down NAF 1.0 in spring 2014

Migration: AFS

- Personal space in NAF 1.0 : /afs/naf.desy.de/user/f/foobar
  - YOU should migrate the space: e.g. using gsiscp
- Personal space in NAF 2.0 : /afs/desy.de/user/f/foobar
  - Quota can be adjusted – starting with 4 GB
- Aditional space: XXL-Volume: : /afs/desy.de/user/f/foobar/xxl
  - Quota can be adjusted – starting with 8 GB
  - Lesser quality of service
- DESY-AFS (NAF 2.0) can be easily accessed from outside:

```
kinit foobar@DESY.de     (enter your password)
aklog –c desy.de
```

- Group space: Admins will take care of this

Migration: dCache

- dCache migration trivial – there is none!

- Access to dCache with the normal tools. Software might need to be adapted to slightly different environment on NAF 2.0
- Data path and access method remain identical

- dcTools and NFS v4.1 dCache mount
  - SL5: dcTools Maintained for short time (until SL6 migration is done)
  - SL6: Working on mounting dCache read-only using NFS 4.1 - make dcTools obsolete
  - Pilot users for NFS 4.1: BELLE running on WGS and Batch (since this morning)

Migration: Lustre

- Lustre will NOT be migrated to NAF 2.0
- Lustre decommissioned since 1.10.2013
- LHCb Lustre space in Zeuthen also to be decomissioned

Migration: Sonas

- Some theory: Sonas is organized in filesets
  - Each person has one fileset per experiment
  - Some physics groups also have (usually larger) filesets
  - A fileset is a management entity within Sonas
- More theory: User ID and Group ID
  - NAF 1.0: “id” uid=1234(foobar) gid=1009(cms) groups=1009(cms)
  - NAF 2.0: “id” uid=5678(foobar) gid=3118 (cms) groups=3118(cms)
  - Same person has two UID/GID pairs … clashes with NFS exports
- Therefore: One fileset can *either* be exported to NAF 1.0 *or* to NAF 2.0
  - Users can decide individually when to change the export
  - Migration steps by expert: Remove fileset from NAF 1.0 – apply UID/GID changes – attach fileset to NAF 2.0 … after ~10 minutes, a Sonas fileset is migrated
  - Group filesets are technically identical, need more organization among users

Current Sonas status:

- ILC: Complete migration to NAF 2.0
- ATLAS & CMS: Some users already migrated to NAF 2.0
  - Some still work in the general-purpose test directories in NAF 2.0. These will disappear at some time!
- User migration: Write an email to naf-helpdesk@desy.de including:
  - NAF 1.0 and NAF 2.0 user name / NAF 1.0 path
  - Date & Time when you want migration
- Quota and management status:
  - NAF 1.0 quota management with certificate authorization
  - NAF 2.0 will need Kerberos authorization – under development
  - “Which user has space in NAF 1.0 and NAF 2.0 Sonas?” – “Is on a respective WGS in NAF 1.0 or NAF 2.0”
  - More/less quota? Write to naf-NNN-support@desy.de
- Currently total Sonas capacity is ~650 TB
  - 200 TB purchased and will be deployed end of the year

Software deployment - CVMFS

- Experiments can still use /afs/desy.de/group for their software
- It has shown however that most software currently resides within CVMFS
- We offer several repositories on NAF 2.0:
  - ATLAS: atlas.cern.ch / atlas-nightlies.cern.ch / atlas-condb.cern.ch
  - CMS: cms.cern.ch
  - General: sft.cern.ch (e.g. for general ROOT version)
  - To come: ilc.desy.de (or the like)
  - … more can be deployed if necessary
- On NAF 2.0 we use the automounter
  - ls –l /cvmfs/ might shown you nothing – perform e.g. cd /cvmfs/cms.cern.ch/ and there you go

Migration: Workgroup-Server

- This is the place where you work interactively
- You can work in parallel with NAF 1.0 and NAF 2.0
  - If you want and/or need to – but not for too long!
  - Your Sonas space is either in NAF 1.0 or in NAF 2.0 … here you have to decide
  - No “big-bang” migration – you can adapt your scripts, test the environment and do the full migration once you are ready
- Basically, NAF 2.0 workgroupserver are there
  - E.g. ssh nafhh-cms01.desy.de
  - Similar to normal DESY workgroupserver
  - SL 5 for the moment – some SL 6 being deployed
  - NAF 1.0 will not have SL 6
  - More WGS to come – and also a load balancer

Migration: Batch Worker Nodes

- NAF 2.0 batch is part of the general purpose DESY BIRD cluster
- BIRD currently consisting of ~1600 cores@SL5 and ~2500 cores@SL6
- NAF 1.0: SGE – NAF 2.0: SonOfGridEngine – quite similar in handling
  - Submission scripts should work with little adaption
  - PROOF on Demand should work on BIRD / NAF 2.0

- NAF 1.0 currently consisting of ~1600 cores
  - December: 1000 cores -> SL6@NAF 2.0 (500 remaining in NAF 1.0)
  - January: 350 cores -> SL6@NAF 2.0 (150 remaining in NAF 1.0)
  - March: 150 cores -> SL6@NAF 2.0 (no batch @ NAF 1.0)
  - April: Shutdown of WGS in NAF 1.0

- Zeuthen currently has ~120 cores in NAF 1.0 batch
  - Moved into “The Zeuthen Farm” or decommissioning (6y old)

… and yet another migration

- Hamburg site currently changing configuration management

- Currently in transition phase to Puppet
  - Will affect ALL SL6 systems in future
  - Will not affect SL5 systems
  - Some systems started with Puppet configuration

- Main changes:
  - RPM list might differ – drop a line to naf-helpdesk@desy.de with what you miss
  - Configuration might be different (e.g. prompt, $PATH …)
  - We will provide simple setup scripts for the environment and needed products

- “ini” is deprecated with SL6 – use “module” instead

What needs to be done?

- Monitoring of all components
  - Some already exists – will need to be made available to users
  - Batch system monitoring and accounting to come
- More documentation

Current documentation and link to registry:

http://it.desy.de/services/computing_infrastructure/national_analysis_facility___naf/index_eng.html

- Fine-Tuning of WGS and Batch configuration
  - E.g. RPMs – usually deployed soon after the request

Support @ NAF 2.0

- Basically the same split support model:
- Fabric issues: (“Login not possible”,”RPM missing”)

  naf-helpdesk@desy.de

  … will pass your question to the adequate second level support if needed
- Experiment issue: (“Code does not compile”,”SONAS quota increase”)

  naf-<VO>-support@desy.de
- Experiments experts (should) know second-level entry points – and can use them if necessary

New and possible further developments

- NX-like service: We use products from StarNet (Live-Client)
  - Intended usecase: Use this tool to connect to a “NAF remote desktop”
  - This “desktop” then allows you to work as on a DESY Desktop
  - Two “NAF remote desktop” server exists: an Ubuntu 12.04 and an SL 6 one.
  - Contact naf-helpdesk if you want to use this product (license needed – offered by DESY)
  - … live demo if time permits

- GPU / MIC / ARM computing … DESY has some GPU & MIC resources. If you need such resources for your work contact us. ARM … in 2014
- HPC Cluster …. DESY has a cluster with fast interconnect. If you need such resources for your work contact us
  - … not for running PROOF or normal MC though ☺
- Several commercial products available at DESY
  - E.g. compilers, … , NeuroBayes☺

NAF 2.0 and schools

Monte Carlo Methods in Advanced Statistics Applications and Data Analysis

- Many schools held in the context of the Terascale Alliance
  - Among them computing schools, tutorials and workshops
- With NAF 2.0 it is extremely easy to host such schools
- Example: Monte-Carlo school in Munich two weeks ago:
  - Setup five WGS for interactive work – batch not needed (but would have been possible)
  - Setup group AFS space for common software
  - Setup 90 school accounts
  - Handout with short login instructions including username/password on the first day of the school
  - Students used their own laptops to ssh to the WGS and work there.
- NAF continues to be available for schools, tutorials and workshops
  - NAF 2.0 just makes things a lot more easy

Summary and Outlook

- NAF 1.0 was (and is) a success
  - But the world has changed – need a redesign to continue success story

- NAF 2.0 is there – you can/should use it! **NOW!**

- NAF 1.0 to be shut down – hopefully soon

Current documentation and link to registry:

http://it.desy.de/services/computing_infrastructure/national_analysis_facility___naf/index_eng.html

ATLAS Wiki: https://wiki-zeuthen.desy.de/ATLAS/WorkBook/NAF2 (public)

CMS Wiki: https://twiki.cern.ch/twiki/bin/view/CMS/HamburgWikiComputingNAF (protected)

ILC Wiki: http://www-flc.desy.de/flc/flcwiki/NAF2Start (public)

- Backup slides

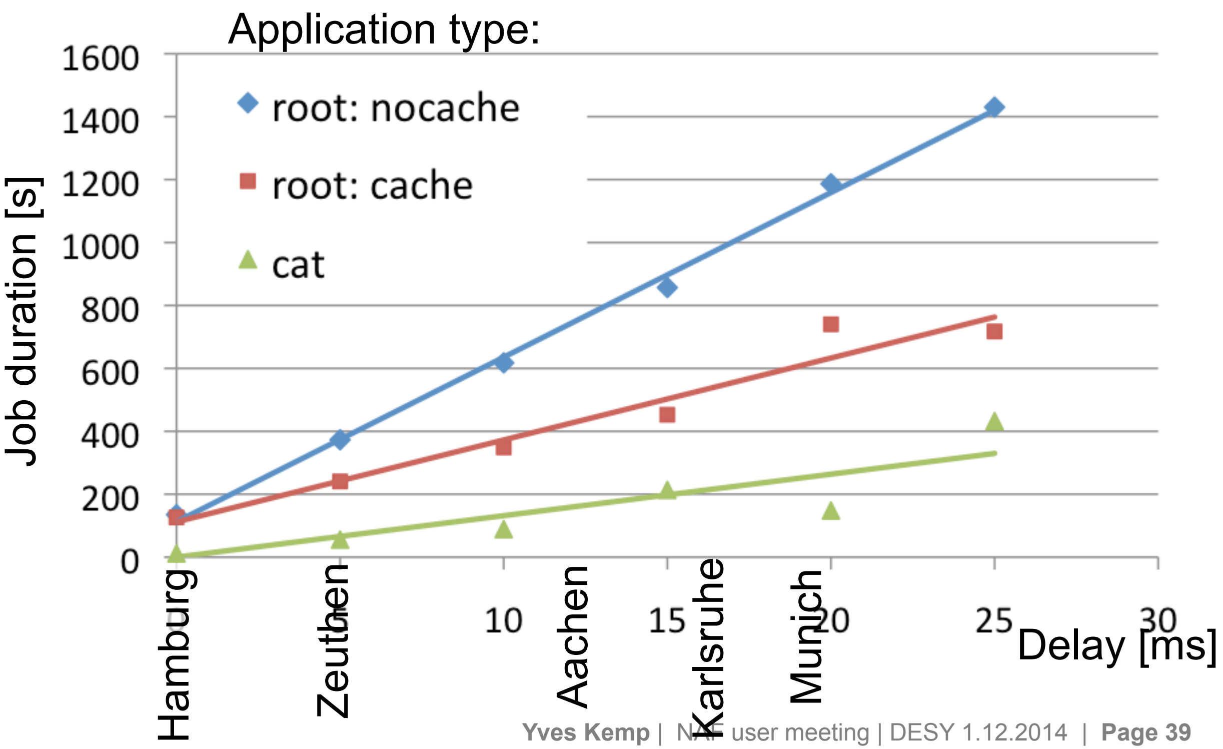
NAF 1.0: Detailed look at networking

- The two-site setup had as an effect that most disturbances at on site had repercussions on the other site and the NAF as a whole
  - Only a very small number of services were deployed in a redundant way
- Investigation on latency effects on dCache NFS 4.1 mounts for ROOT file access: Lab emulation of latency

Reading files over WAN using dCache NFS v4.1

Ping times:
HH-WN -> HH-dCache
<0.2 ms

HH-WN -> ZN-dCache
~ 5.5. ms



Sonas NFS mount and internal / external users

- DESY users primary group: cms
- External users primary group: af-cms

- By default: files in Sonas readable by all CMS people:
  - Solution: DESY are given additional secondary group af-cms

- Make sure that primary group when on NAF 2.0 is af-cms
  - Need to change this on WGS and batch for internal people
  - Performed using a script running at login time on WGS
  - Batch honors primary group at submission time

- People still can decide differently: e.g. chown …