The National Analysis Facility @ DESY

Yves Kemp for the NAF team

DESY IT Hamburg & DV Zeuthen

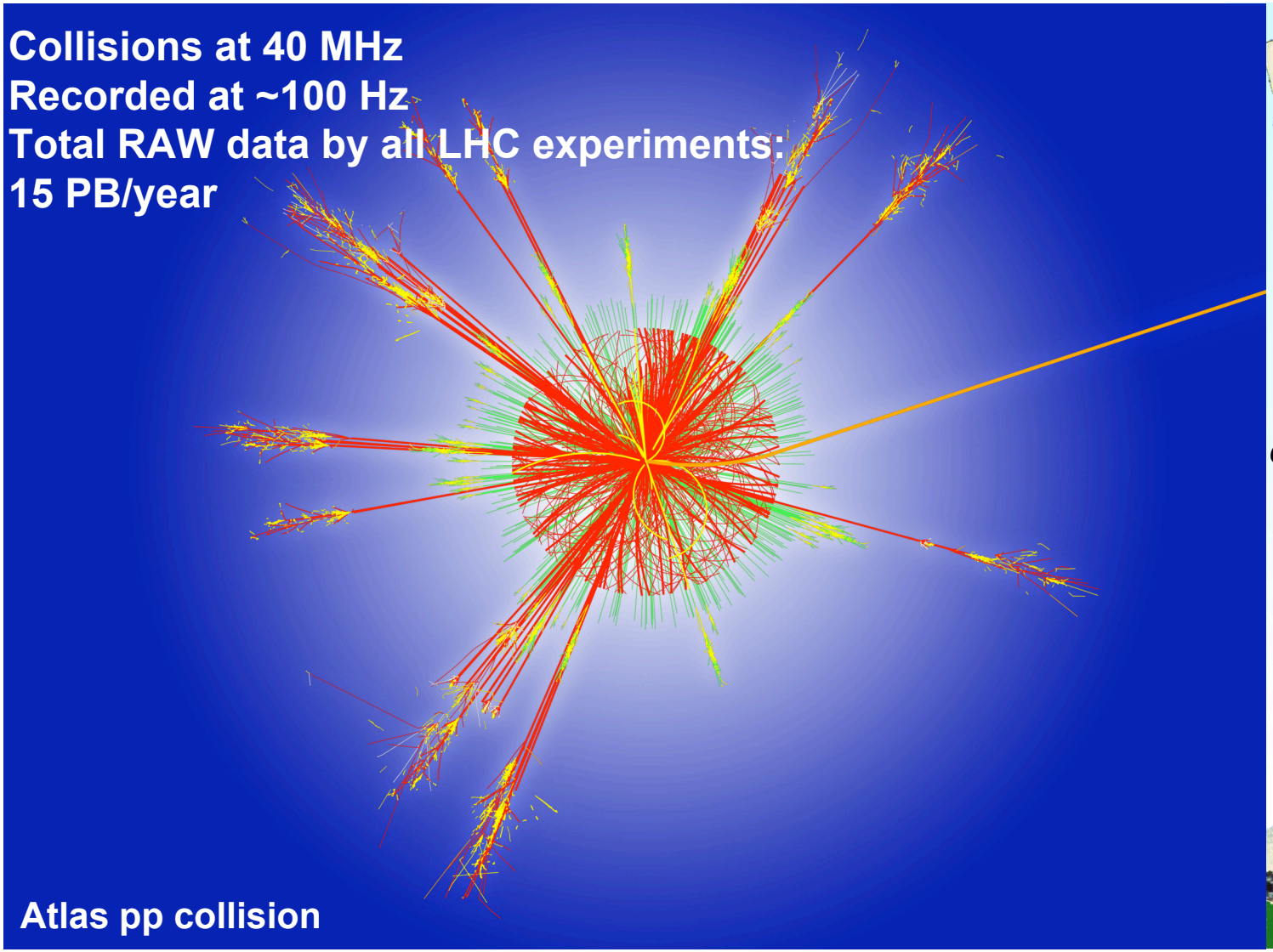
10.9.2008

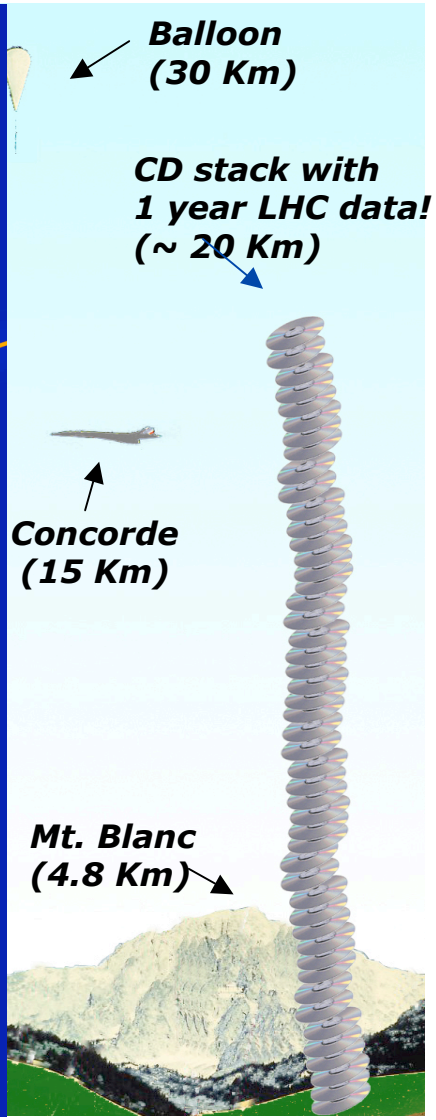
GridKA School

- **NAF: National Analysis Facility**
  - **Why a talk about an “Analysis Facility” at a Grid School?**

- **This talk will show you**
  - **what kind of requirements for computing resources Analysis has**
  - **where the Grid can meet them and where it cannot**
  - **planning details of the NAF**

HEP Computing: Data centric





Different tasks: Different requirements

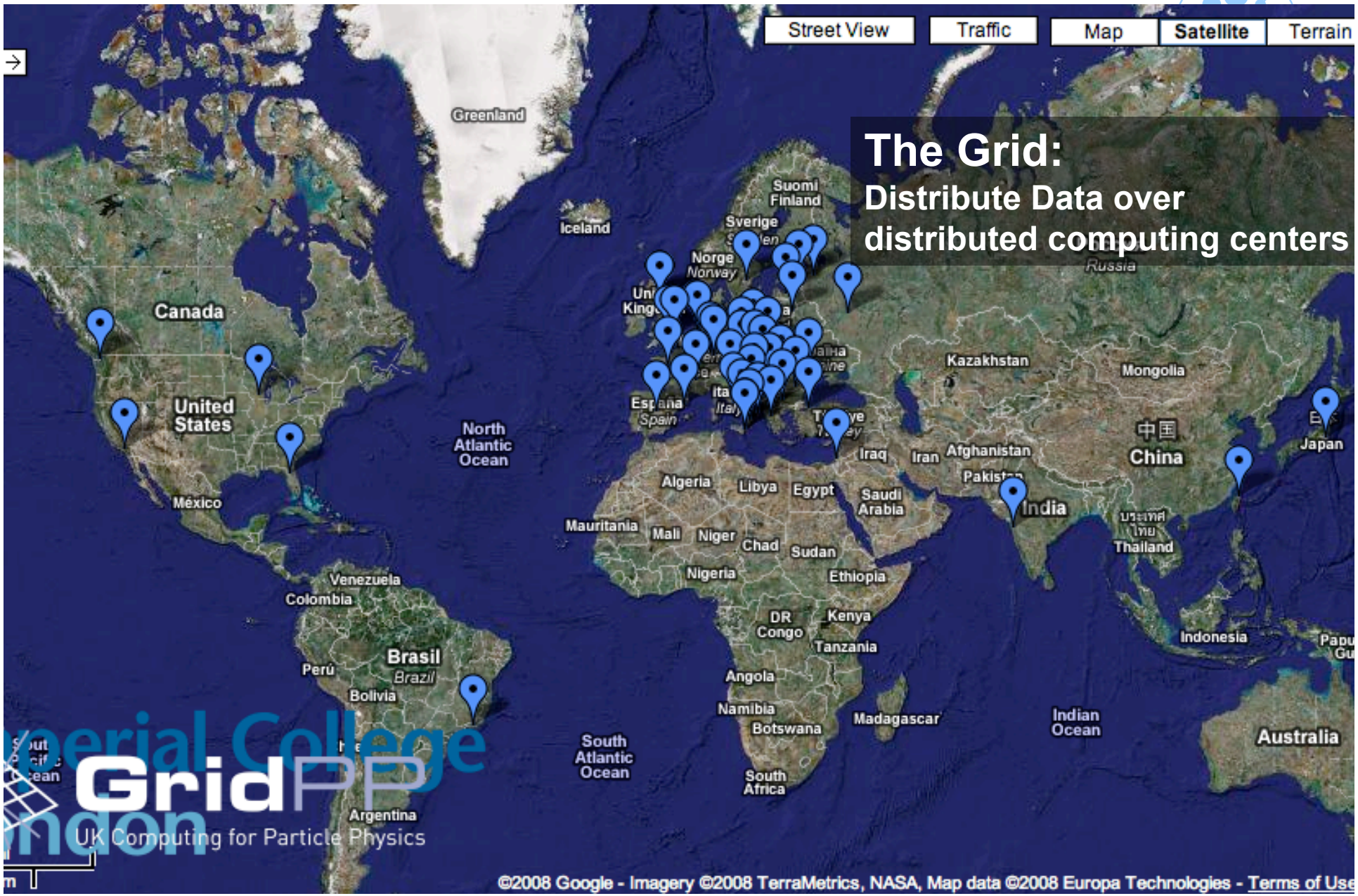
HAMBURG • ZEUTHEN

Coordinated & global tasks

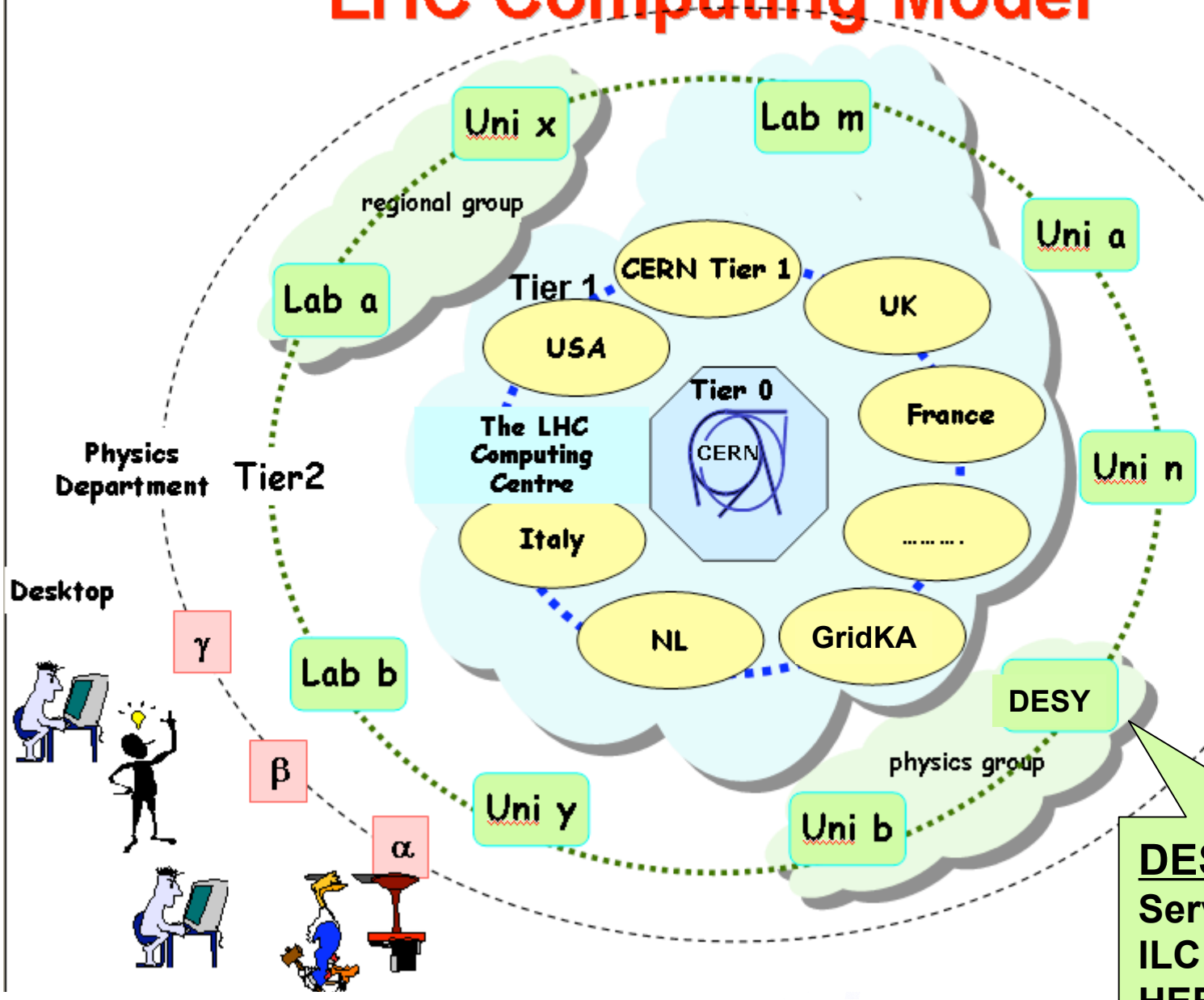
- **MC Production**
  - Event Generation: no I; small O; little CPU
  - Detector Simulation: small I; large O & CPU
- **Event Reconstruction/Reprocessing**
  - Reprocessing: full I; full O; large CPU
  - Selections: large I; large O; large CPU

Uncoordinated, unstructured & local tasks

- **Analysis**
  - Usually: large I; small O; little CPU
  - Performed by many users, many times!
  - LHC StartUp phase: Short turn-around


Street View
Traffic
Map
Satellite
Terrain
The Grid:
Distribute Data over
distributed computing centers
Greenland
Iceland
Suomi
Finland
Sverige
Norge
Norway
Russia
Canada
United
States
México
North
Atlantic
Ocean
Kazakhstan
Mongolia
中国
China
Japan
Iraq
Iran
Afghanistan
Pakistan
India
Algeria
Libya
Egypt
Saudi
Arabia
Mauritania
Mali
Niger
Chad
Sudan
Nigeria
Ethiopia
Kenya
DR
Congo
Tanzania
Angola
Namibia
Botswana
Madagascar
South
Atlantic
Ocean
South
Africa
Indian
Ocean
Indonesia
Australia
Thailand
Venezuela
Colombia
Brasil
Brazil
Perú
Bolivia
Argentina
GridPP
UK Computing for Particle Physics
©2008 Google - Imagery ©2008 TerraMetrics, NASA, Map data ©2008 Europa Technologies - Terms of Use

LHC Computing Model



Each layer:

Specialized for certain tasks

e.g. T2:

-Analysis

-User access

-AOD storage

DESY Grid:
Serves Atlas&CMS
ILC & Calice
HERA-Experiments
…

Do we need something in addition?

- Grid and the Tier model well suited for
  - Global & coordinated tasks
- Analysis
  - Local & uncoordinated, unstructured
- Provide best possible infrastructure and tools for German researchers
  - In addition to global Grid resources
- Join forces and create synergies among German scientists

- The NAF: National Analysis Facility
  - Located at DESY: Data is there

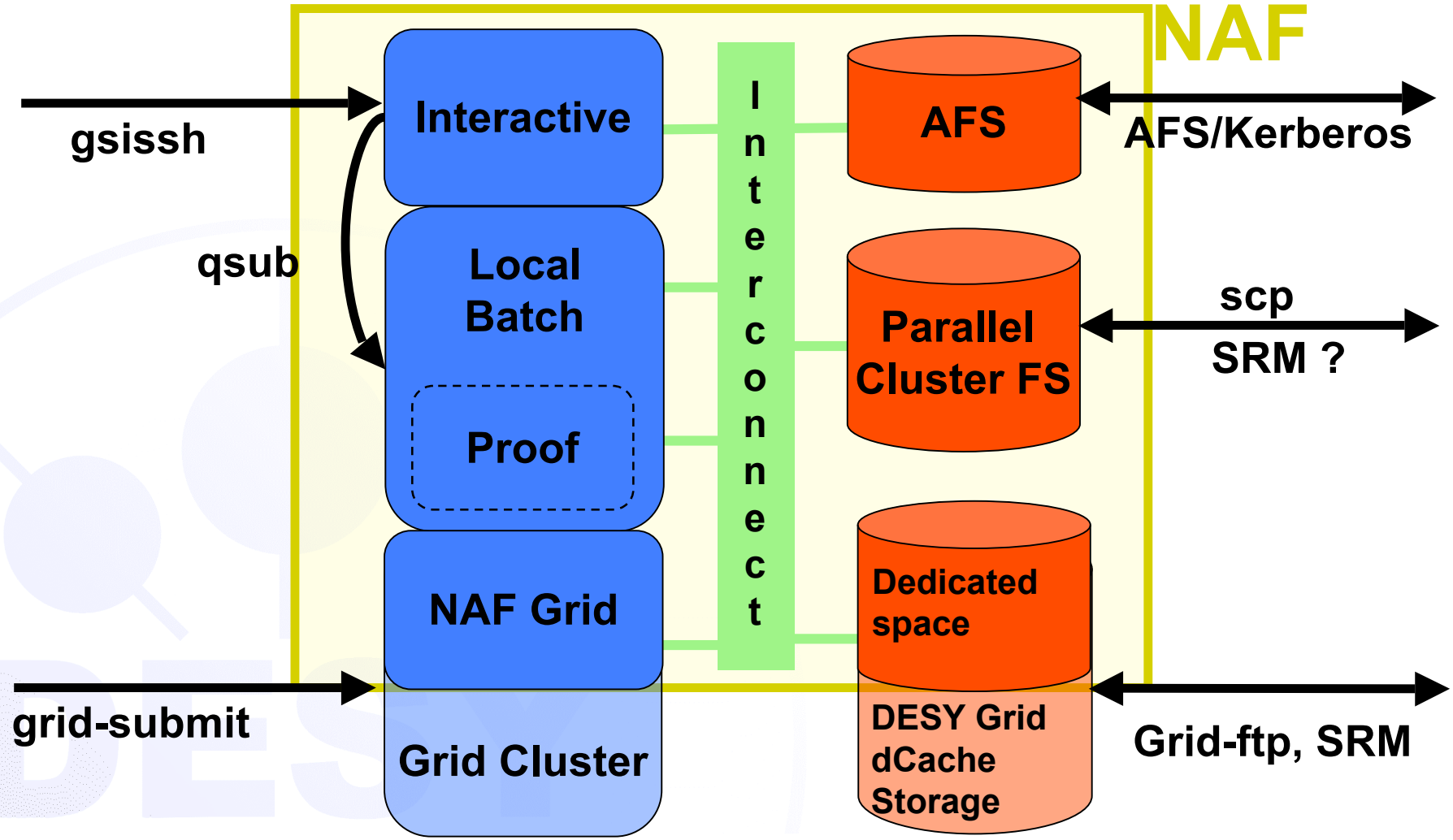
The frame for the NAF:

- The NAF is part of the Strategic Helmholtz Alliance
  - More: http://terascale.desy.de/
- Only accessible by German research groups for LHC and ILC tasks
  - Planned for a size of about 1.5 av. Tier 2, but with more data
  - Starting as joint activity @ DESY
- Requirements papers from German Atlas and CMS groups

Starting with Atlas & CMS

- **Requirement papers. Some points:**
- **Interactive login**
  - **Code development & testing, Experiment SW and tools**
  - **Uniform access**
  - **Central registry**
- **Personal/group storage**
  - **AFS home directories (and access to other AFS cells)**
- **High-capacity /High-bandwidth storage**
  - **Local part (potentially with backup)**
  - **Grid part: Enlargement of the T2 part**
- **Batch-like resources:**
  - **Local access: short queue, for testing purpose**
  - **Large part (only) available via Grid-mechanisms**
  - **Fast response wanted for local&Grid**
- **Hosted Data:**
  - **AODs (Full set in case for Atlas, maybe trade some for ESD?)**
  - **TAG database**
  - **User/Group data**
- **Additional services**
  - **PROOF farm, with connection to high-bandwidth storage**
- **Flexible setup**
  - **Allows reassignment of hosts between different types of services**

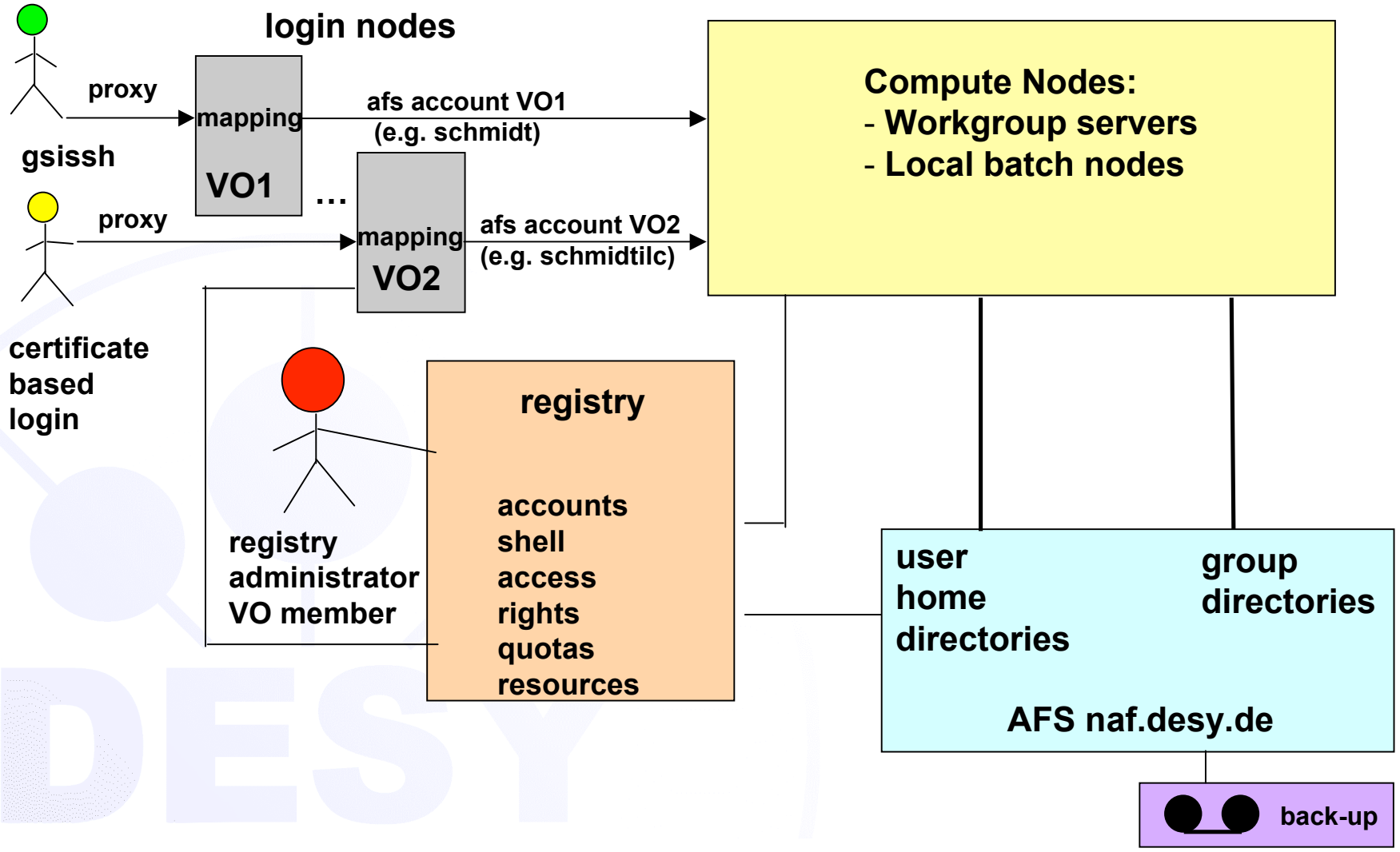
Infrastructure building blocks



Grid Part of NAF

- Use VOMS!! voms-proxy-init --voms
  - atlas cms
  - atlas:/atlas/de cms:/cms/dcms
- NAF Grid ressources integrated into DESY Grid Cluster
  - Separate Fairshare and Priority for German users
  - Access to storage based on VOMS groups/roles to come!

NAF login, interactive



IO and Storage

- New AFS cell: naf.desy.de
  - User & Working group directories
  - Special software area
  - Safe and distributed storage

- Cluster File System
  - High Bandwidth (O(GB/s)) to large Storage (O(10TB)
  - Copy data from Grid, process data, save results to AFS or Grid
  - “Scratch-like” space, lifetime t.b.d., but longer than typical job
  - Locally connected via InfiniBand, remote access via TCP/IP

- dCache
  - Well-known product and access methods
  - Central entry point for data import and exchange
  - Special space for German users

Storage organization

- ATLAS
  - "DESY has 100% of the AODs" : Distributed between HH and Zn
  - More than the nominal T2 pledge: Additions are "the NAF part"
  - RDO/ESD at a smaller level, if requested and if space
- CMS
  - Concept of "T2 hosting an analysis"
    - DESY-HH try to host as many analysis as possible
  - Have all interesting data for physics
- ILC/CALICE
  - Already have MC data (ILC) and real data (CALICE)
  - But at a smaller scale
- dCache SE to host these data!

Access to storage

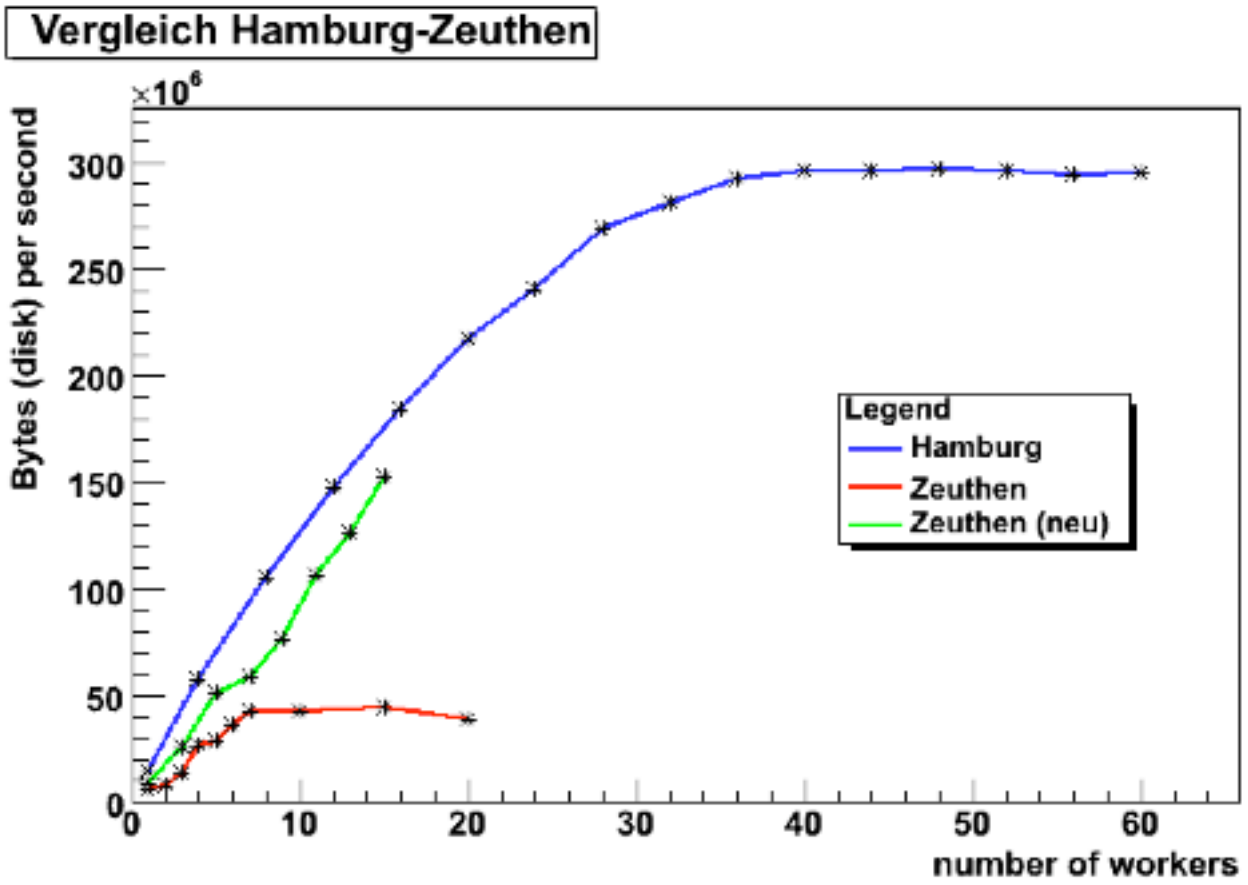
- AFS well known product, access clear
- Lustre: is a cluster filesystem
  - Use as a normal filesystem
  - (OK: some limitations concerning locking and handling of many small files…)
- dCache: different access methods:
  - Via Grid methods: LFC …
  - /pnfs mount: BUT: Security and performance problems!
  - DESY summer student Malte Nuhn: Development of secure and low-resource consuming tools for replacing /pnfs mount

Software

- **Experiment specific software: Grid and Interactive world:**
  - **DESY provides space and tools: Experiments install their software themselves**
  - **Because of current nature of Grid and Interactive parts: Two different areas**
- **Common software:**
  - **Grid world: Standard worker node installation**
  - **Interactive world: Compilers, debuggers… ROOT, CERNLIB**
- **Operation System:**
  - **Currently all Grid WNs on SL4 (64 bit)**
  - **InteractiveSL4 (64bit) (some SL5 testing machines). No SL3**

PROOF: Experience from CMS

Test by UniHH running proof under SGE batch, data on Lustre FS.



Infiniband

tcp buffer tuning

no tcp buffer tuning

Courtesy of Wolf Behrenhoff

PROOF from CMS, cntd.

- **running on SGE batch farm**
  - **access to Lustre, dCache, ...**
- **every user starts his/her own PROOF cluster**
  - **crashes, segfaults, ... never affect others**
  - **no need to deal with authentication, permissions, ...**
  - **simple setup, scripted start/stop**
  - **no version/compatibility problems. One user can run 5.14 (CMSSW 1.x), others can use 5.18 at the same time**
- **First user doing real analysis since July 4th**

Courtesy of Wolf Behrenhoff

More information: See PROOF/ROOT tutorial for general PROOF or attend the CMS course this afternoon

Support, Documentation, NUC

- **Docu**
  - Main entry point: http://naf.desy.de/
  - Links to experiment-specific pages linked from here
- **Support**
  - General entry point: naf-helpdesk@desy.de
  - Experiment-specific support: See their docu
- **NAF Users Committee NUC**
  - Atlas: Wolfgang Ehrenfeld, DESY Jan-Erik Sundermann, Freiburg
  - CMS: Hartmut Stadie, Uni-HH Carsten Hof, Aachen

Summary & Outlook

- NAF already has many active users
- All building blocks in place
  - Still tuning needed for some
- Additional services wanted
  - e.g. TAG-DB for ATLAS: to come
- YOU should get an account (if you not already have one:-))
  - CMS tutorial this afternoon on NAF
  - ATLAS tutorial next week @ Munich on NAF

Backup:
Current and prospected hardware

- NAF-Batch: currently 264 cores (HH+Zn)
  - 2008: HH +256 ; Zn +128
- NAF-Grid: German groups have each:
  - 10% of 1262 cores fairshare.
- Lustre: ~60 TB (in HH)
- dCache: (T2 & NAF !!)
  - Enlargement of HH 480 TB / ZN 90TB in 2008
- + other backbone systems