A National Analysis Facility @ DESY

Yves Kemp for the NAF team
DESY IT Hamburg & DV Zeuthen

1.4.2008
65. Physics Research Committee (PRC) - Open Session
DESY Hamburg
Collisions at 40 MHz
Recorded at ~100 Hz
Total RAW data by all LHC experiments:
15 PB/year

Atlas pp collision

Balloon (30 Km)
CD stack with 1 year LHC (~ 20 Km)
Concorde (15 Km)
Mt. Blanc (4.8 Km)
Different tasks: Different requirements

- **MC Production**
  - Event Generation: no I; small O; little CPU
  - Detector Simulation: small I; large O & CP

- **Event Reconstruction/Reprocessing**
  - Reprocessing: full I; full O; large CPU
  - Selections: large I; large O; large CPU

- **Analysis**
  - Usually: large I; small O; little CPU
  - Performed by many users, many times!
  - LHC StartUp phase: Short turn-around
The Grid: Distribute Data over distributed computing centers.
Each layer is specialised for certain tasks

- 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
    - Grid & local (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, may 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

AFS
AFS/Kerberos

scp
SRM?

grid-ftp, SRM

grid-submit

gsissh

Interactive

Local Batch

Proof

NAF Grid

Parallel Cluster FS

Dedicated space

DESY Grid dCache Storage

NAF

Connecting

PROOF

Grid Cluster

PRC DESY 1.4.2008

LHC Computing: NAF @ DESY

Yves Kemp
Infrastructure building blocks

Interactive
Local Batch
Proof
NAF Grid
Grid Cluster

AFS
Parallel Cluster FS
Dedicated space
DESY Grid dCache Storage

NAF

DESY
Infrastructure building blocks

Interactive

Local Batch

Proof

NAF Grid

AFS

Parallel Cluster FS

Dedicated space

READY Tier2!

Grid Cluster

DESY Grid dCache Storage

PRC DESY 1.4.2008

LHC Computing: NAF @ DESY

Yves Kemp
Infrastructure building blocks

- Interactive
- Local Batch
- Proof
- NAF Grid
- Grid Cluster
- AFS
- Parallel Cluster FS
- Dedicated space
- DESY Grid dCache Storage

Ready User testing

Ready In Use

Ready Tier2!
Infrastructure building blocks

- Ready User testing
  - Interactive
  - Local Batch
  - Proof
  - NAF Grid

- Ready In Use
  - Grid Cluster

- Ready Tier2!
  - Dedicated space
  - Parallel Cluster FS
  - DESY Grid dCache Storage

- NAF
  - AFS
  - Test instance soon
Infrastructure building blocks

Interactive
Local Batch
Proof
NAF Grid
Grid Cluster
AFS
Parallel Cluster FS
Dedicated space
DESY Grid dCache Storage

Ready
User testing

Coordinating with users

Ready In Use

Ready Tier2!

Test instance soon
Summary & Outlook

- DESY well established in the LHC and ILC computing

- We need a National Analysis Facility for German LHC and ILC groups

- Key parts are in place
  - Alpha test ongoing with local users
  - First external users mid April
  - Close cooperation with experiments
Backup Slides
The first hardware

- Hardware Computing:
  - 6 x 16 DualCPU-Quadcore Blades (HP-Proliant BL460c)
  - 2GB RAM/core, 146 GB HD/Blade
  - Infiniband HCA
- Hardware Storage:
  - 7 x SUN thumpers (17.5TB/box at raid 6) for dCache pools & Lustre
  - 8 x DELL Poweredge 2950 with 8x146 GB SAS Disks for infrastructure and AFS
- Other hardware: Racks, Infiniband Switch, Infrastructure servers, ...

- Hardware assignment flexible:
  - Following needs of experiments
NAF login, interactive

- Login nodes
  - proxy
  - gsssh
  - certificate based login
    - registry administrator VO member
      - registry
        - accounts
        - shell
        - access
        - rights
        - quotas
        - resources
  - mapping
    - VO1
      - afs account VO1 (e.g. schmidt)
  - mapping
    - VO2
      - afs account VO2 (e.g. schmidt@ic)

- Compute nodes
  - user home directories
    - AFS naf.desy.de
  - group directories
  - back-up
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

- dCache
  - Well-known product and access methods
  - Central entry point for data import and exchange
  - Special space for German users
First peek @ Login & AFS
**Batch @ NAF**

- **Grid Batch**
  - Integrated in Desy Grid Infrastructure (grid-ce3.desy.de)
  - Dedicated Fairshare with higher priority
  - Access using VOMS proxies

- **Local Batch**

  ![Diagram of Local Batch System]

  - Interactive Workgroup server
  - qsub
  - SGE Batchsystem
  - WN
  - scratch
  - WN
  - scratch
  - WN
  - scratch
  - dC

---

PRC DESY 1.4.2008

LHC Computing: NAF @ DESY

Yves Kemp 16
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:
    - Workgroup server installation: Compilers, debuggers...
    - No Browser, Mailclient, ....
    - ROOT, CERNLIB?? (Are the ones shipped with the experiment frameworks OK?)

- Operation System:
  - Currently all Grid WNs on SL4
  - InteractiveSL4 (64bit) (maybe some SL5). No SL3
Support

- Technical aspects:
  - DESY can provide technical tools like mailing lists, request tracker, wiki, hypernews... if needed
  - GGUS might be integrated

- Organisational aspects
  - The experiments MUST provide first level support
    - Filter user questions
    - Transmit fabric issues to NAF admins
  - DESY will provide second level support

- We NAF operators need fast feedback: NAF Users Board
  - E.g. two experienced users from each experiment for technical advisory and fast feedback + 2 NAF operators
First user experience: Feedback

- Main users using the NAF Grid share:
  - Adrian Vogel (ILC): ~8k h, 10k Jobs
    - Machine-induced background studies, full Geant4 detector simulation
  - Manuel Giffels (CMS): ~70 h, 100 Jobs
    - Private background production (bb- >2mu)
  - Walter Bender, Daiske Tornier (CMS):
    - Exotica, private Alpgen production

- Experience:
  - VOMS group works fine, no problems