Openlab Status and Plans 2003/2004

****

Openlab - FM Workshop 8 July 2003

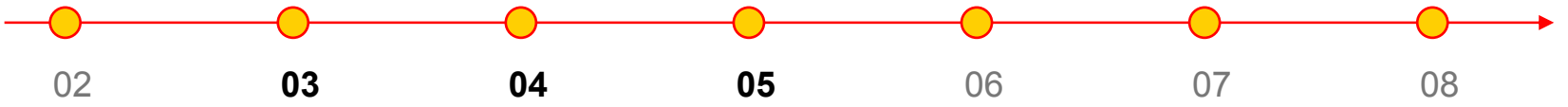
CERN
openlab for DataGrid applications

CERN openlab

LCG

CERN Openlab

- **Framework for industrial collaboration**
- **Evaluation, integration, optimization**
  - of cutting-edge technologies
  - Without the constant "pressure" of a production service
- **3 year lifetime**



openlab: A technology focus

- **Industrial Collaboration**
  - **Enterasys, HP, and Intel were our partners in Q1**
  - **IBM joined in Q2:**
    - **Storage subsystem**
  - **Technology aimed at the LHC era**
    - **Network switches at 10 Gigabits**
    - **Rack-mounted servers**
    - **64-bit Itanium-2 processors**
    - **StorageTank**

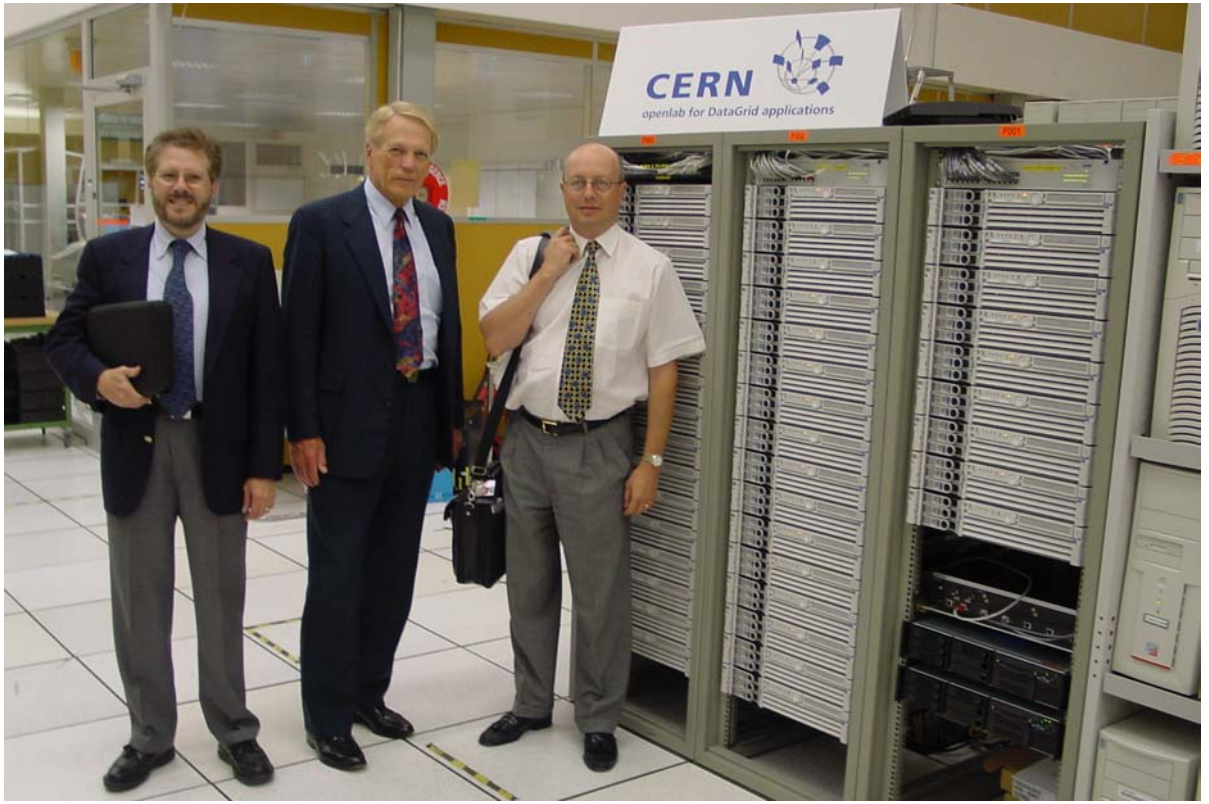
CERN
openlab for DataGrid applications

Main areas of focus

- The cluster
- The network
- The storage system
- Gridification
- Workshops

CERN
openlab for DataGrid applications

The cluster

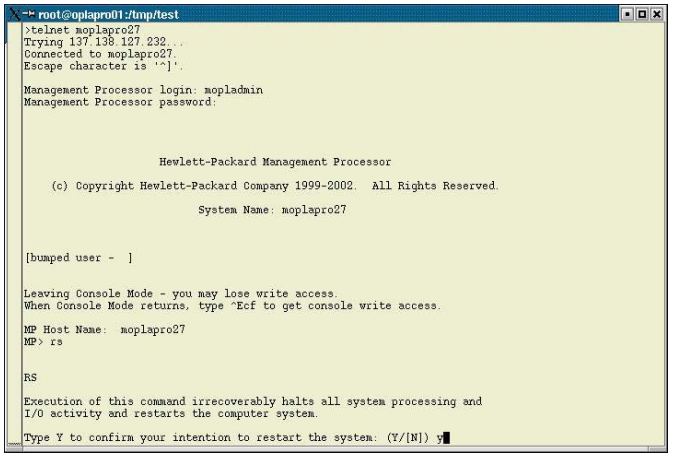


opencluster in detail

- **Software integration:**
  - 32 nodes + development nodes
  - Fully automated kick-start installation
  - Red Hat Advanced Workstation 2.1
  - OpenAFS 1.2.7, LSF 5.1
  - GNU, Intel, ORC Compilers
    - ORC (Open Research Compiler, used to belong to SGI)
  - CERN middleware: Castor data mgmt
  - CERN Applications
    - Porting, Benchmarking, Performance improvements
  - Database software (MySQL, Oracle)
    - Not yet

Remote management

- **Built-in management processor**
  - **Accessible via serial port or Ethernet interface**
- **Full control via panel**
  - **Reboot**
  - **power on/off**
  - **Kernel selection (future)**



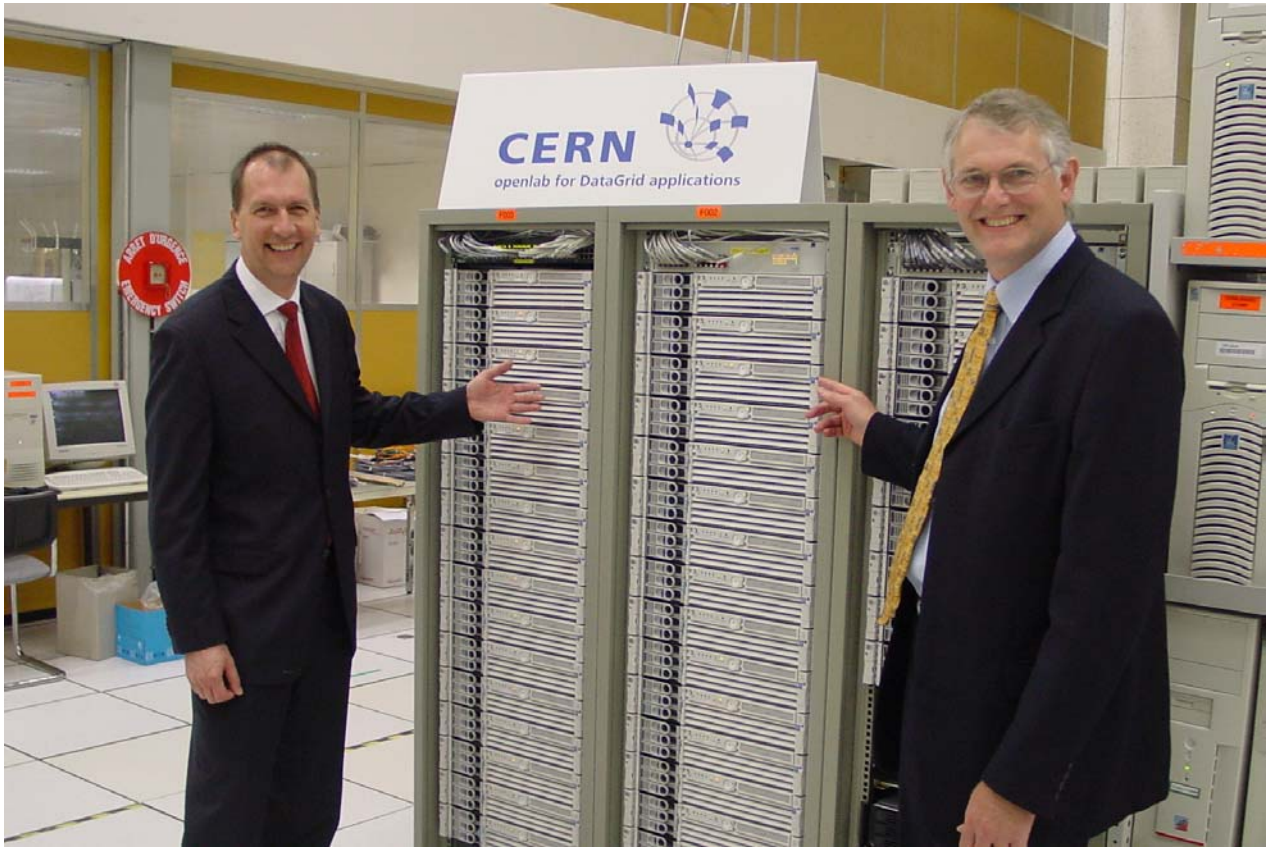
opencluster

- **Current planning:**
  - Cluster evolution:
    - 2003: 64 nodes ("Madison" processors @ 1.5 GHz)
      - Two more racks
    - 2004: Possibly 128 nodes, Madison++ processors)
  - Redo all relevant tests
    - Network challenges
    - Compiler updates
    - Application benchmarks
    - Scalability tests
  - Other items
    - Infiniband tests
    - Serial-ATA disks w/RAID

Make the cluster available to all relevant LHC Data Challenges

CERN
openlab for DataGrid applications

64-bit applications



Program porting status

- **Ported:**
  - **Castor (data management subsystem)**
    - **GPL. Certified by authors.**
  - **ROOT (C++ data analysis framework)**
    - **Own license. Binaries both via gcc and ecc. Certified by authors.**
  - **CLHEP (class library for HEP)**
    - **GPL. Certified by maintainers.**
  - **GEANT4 (C++ Detector simulation toolkit)**
    - **Own license. Certified by authors.**
  - **CERNLIB (all of CERN's FORTRAN software)**
    - **GPL. In test.**
      - **Zebra memory banks are I*4**
  - **ALIROOT (entire ALICE framework)**
- **Not yet ported:**
  - **Datagrid (EDG) software**
    - **GPL-like license.**

Benchmark: Rootmarks/C++

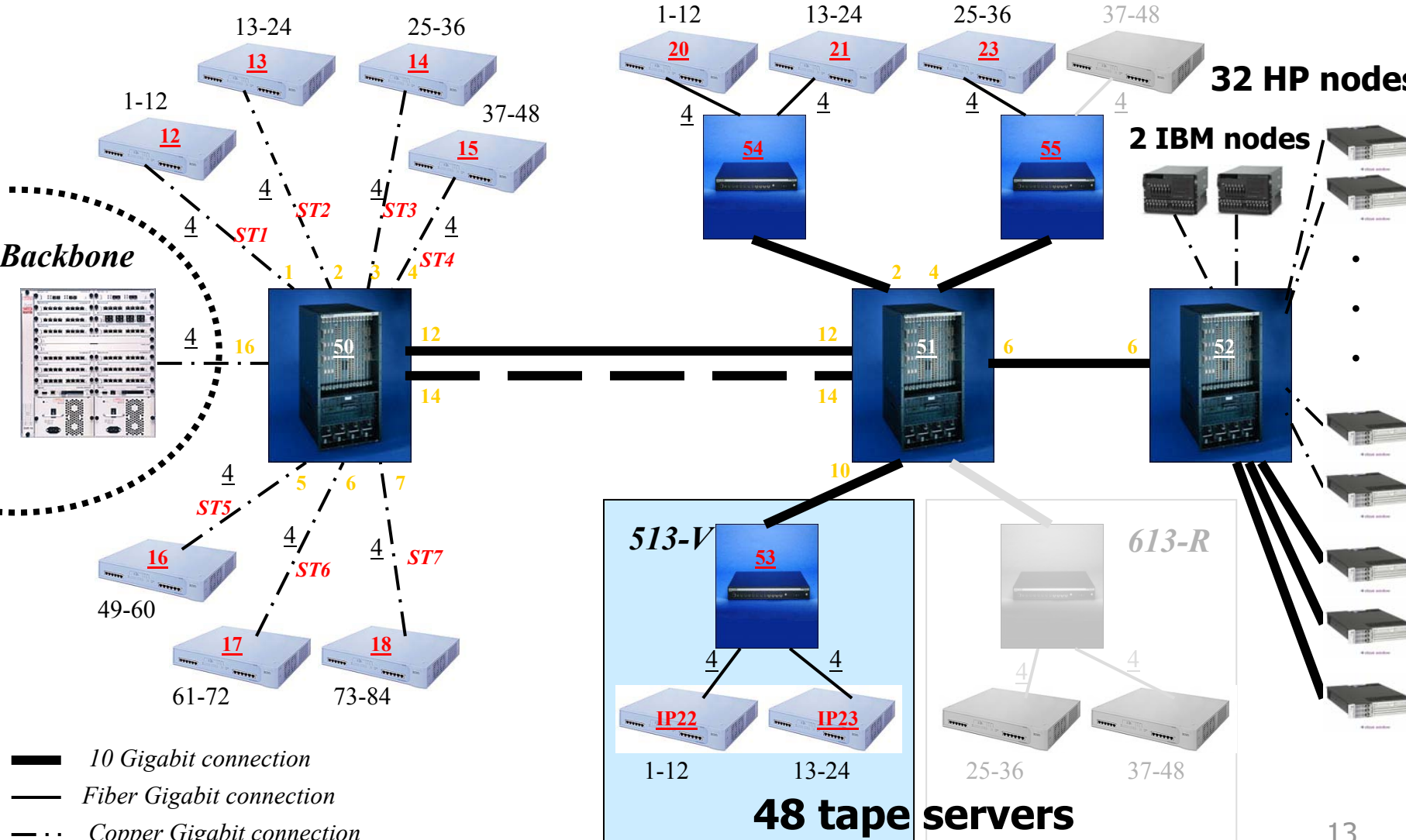
All jobs run in "batch" mode ROOT 3.05.03	Itanium 2 @ 1000MHz (gcc 3.2, O3)	Itanium 2 @ 1000MHz (ecc7 prod, O2)	Itanium 2 @ 1000MHz (ecc7 prod,O2, ipo,prof_use)	Expectations for Madison (1500 MHz) with ecc8
stress –b -q	437	499	585	900++
bench –b -q	449	533	573	900++
root -b benchmarks.C -q	335	308	360	600++
Geometric Mean	404	434	494	

René's own 2.4 GHz P4 is normalized to 600 RM with gcc.

The network

CERN
openlab for DataGrid applications

Enterasys 2Q 2003



10GbE: Back-to-back tests

- **3 sets of results (in MB/s):**

10 km fibres

No tuning, →	1 stream	4 streams	12 streams
1500B	127	375	523
9000B	173	364	698

+ kernel tuning	1 stream	4 streams	12 streams
1500B	203	415	497
9000B	329	604	662

+ driver tuning	1 stream	4 streams	12 streams
1500B	275	331	295
9000B	693	685	643
16114B	755	749	698

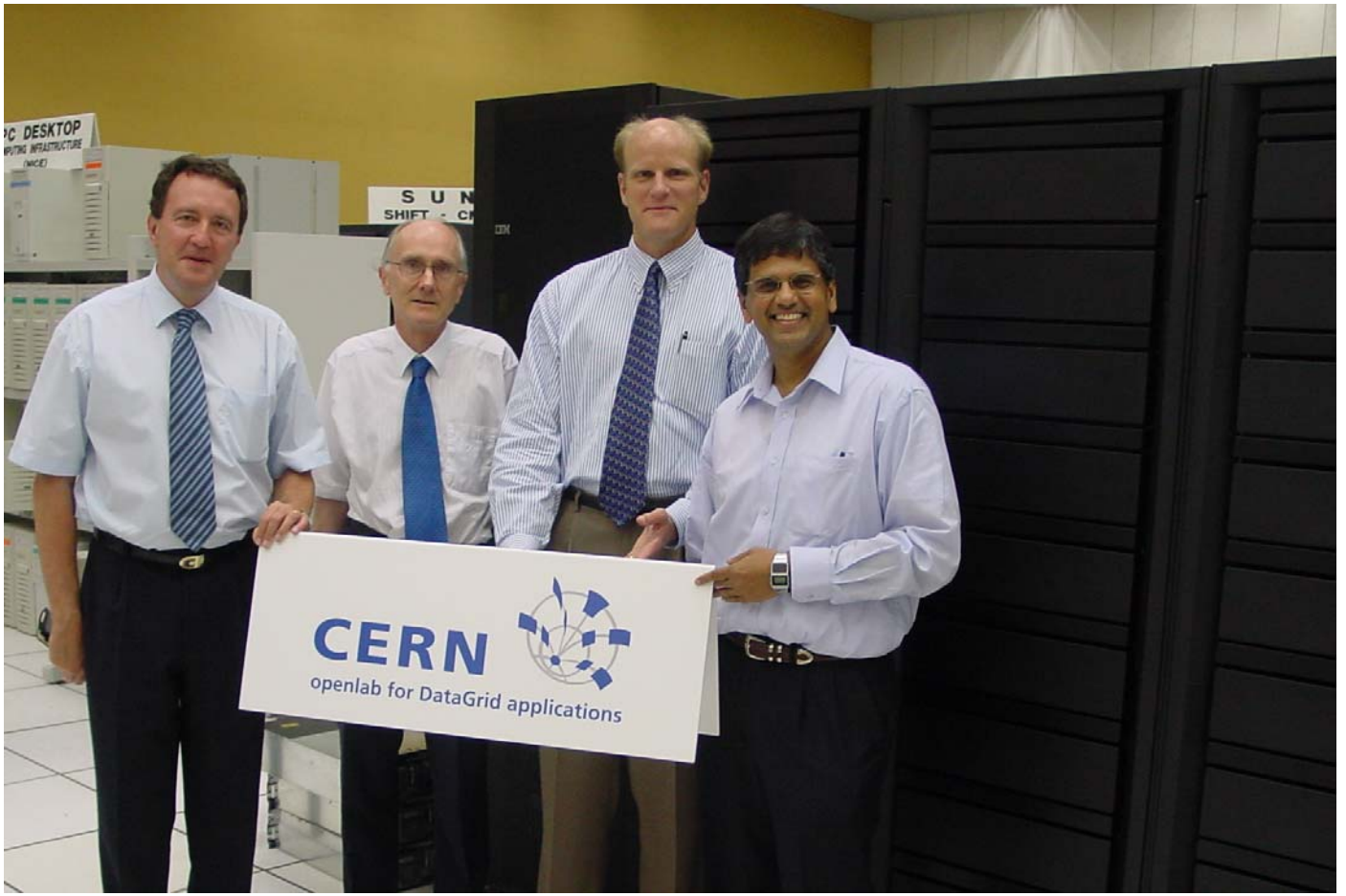
Summer student to work on measurements: Glenn June 2003

Saturation of PCI-X around 800-850 MB/s

Disk speed tests

- **Various options available:**
  - **3 internal SCSI disks:**
    - **3 x 50 MB/s**
  - **Intel PCI RAID card w/S-ATA disks**
    - **4 x 40 MB/s**
  - **Total:**
    - **310 MB/s**
  - **Our aim:**
    - **Reach 500++ MB/s**
    - **Strategy:**
      - **Deploy next-generation PCI-X 3ware 9500-16/-32 RAID card**

The storage system



Initial StorageTank plans

- **Installation and training: Done**

Summer student to work on measurements: Bardur

- **Establish a set of standard performance marks**
  - **raw disk speed**
  - **disk speed through iSCSI**
  - **file transfer speed through iSCSI & Storage Tank**
- **Storage Tank file system initial usage tests**
- **Storage Tank replacing Castor disk servers ?**
  - **Tape servers reading/writing directly from/to Storage Tank file system**

Further ST plans

- **Openlab goals include:**
  - **Configure ST clients as NFS servers**
    - **For further export of data**
  - **Enable GridFTP access from ST clients**
    - **Make ST available throughout a Globus-based Grid**
  - **Make available data that is currently stored in other sources**
    - **through Storage Tank as part of a single name space.**
  - **Increase the capacity: 30 TB → 100 TB → 1000 TB**

Gridification

Opencluster and the Grid

- Globus 2.4 installed
  - Native 64 bit version
  - First tests with Globus + LSF have begun
- Investigation of EDG 2.0 software started
- Joint project with CMS
  - Integrate opencluster alongside EDG testbed
  - Porting, Verification
    - Relevant software packages (hundreds of RPMs)
    - Understand chain of prerequisites
    - Exploit possibility to leave control node as IA-32
- Interoperability with EDG/LCG-1 testbeds
- Integration into existing authentication and virtual organization schemes
- GRID benchmarks
    - To be defined
    - Certain scalability tests already in existence

PhD student to work on Grid porting and testing: Stepher

Workshops

CERN
openlab for DataGrid applications

Storage Workshop

- Data and Storage Mgmt Workshop
- March 17th – 18th 2003
- Organized by the CERN openlab for Datagrid applications and the LCG

- Aim: Understand how to create synergy between our industrial partners and LHC Computing in the area of storage management and data access.

- Day 1 (IT Amphitheatre)

- Introductory talks:
- 09:00 – 09:15 Welcom
- 09:15 – 09:35 Openla
- 09:35 – 10:15 Gridif ent shortcomings (Kunszt)
- 10:15 – 11:15 Coffe

- The current situati
- 11:15 – 11:35 The
- 11:35 – 12:05 CA
- 12:05 – 12:25 IDE Disk Se leinhard)
- 12:25 – 14:00 Lunch

- Preparing for the future
- 14:00 – 14:30 ALICE Data Challenges: On the way to recordin s (Divià)
- 14:30 – 15:00 Lessons learnt from managing data in the European Data Grid (Kunszt)
- 15:00 – 15:30 Could Oracle become a player in the physics data management? (Shiers)
- 15:30 – 16:00 CASTOR: possible evolution into the LHC era (Barring)
- 16:00 – 16:30 POOL: LHC data Persistency (Düllmann)
- 16:30 – 17:00 Coffee break
- 17:00 – Discussions and conclusion of day 1 (All)

- Day 2 (IT Amphitheatre)
- Vendor interventions; One-on-one discussions with CERN

DONE

CERN
openlab for DataGrid applications

2nd Workshop: Fabric Management

- **Fabric Mgmt Workshop (Final)**
- **July 8th – 9th 2003 (Sverre Jarp)**
- **Organized by the CERN openlab for Datagrid applications**

- **Aim: Understand how to create synergy between our industrial partners and LHC Computing in the area of fabric management. The CERN talks will cover both the Computer Centre (Bld. 513) and one of the LHC online farms, namely CMS.**

- **External participatio**
- **HP: John Manley, M** **ul Murray, Fernando Pedone, Peter Toft**
- **IBM: Brian Carpent** **ichard Ferri, Kevin Gildea, Michel Roethlisberger**
- **Intel: Herbert Cor**

TODAY

- Day 1 (IT Amphithe
- **Introductory tal**
- **09:00 – 09:15 W**
- **09:15 – 09:45 Intro** **Jarp)**
- **09:45 – 10:15 Setting the sc** **1 Centres at CERN (T. Cass)**
- **10:15 – 10:45 Coffee break**
- **Part 2:**
- **10:45 – 11:15 Setting the scene (2): Plans for control and mon** **LHC online farm (E.Meschi/CMS)**
- **11:15 – 12:00 Concepts: Towards Automation of computer fabrics (M. Barroso-Lopez)**
- **12:00 – 13:30 Lunch**
- Part 3
- **13:30 – 14:00 Deployment (1): Maintaining Large Linux Clusters at CERN (T. Smith)**
- **14:00 – 14:30 Deployment (2): Monitoring and Fault tolerance (H. Meinhard)**
- **14:30 – 15:00 Physical Infrastructure issues in a large Centre (T. Cass)**
- **15:00 – 15:30 Infrastructure issues for an LHC online farm (A. Racz)**
- **16:00 – 16:30 Coffee break**

CERN
openlab for DataGrid applications

2nd Workshop: Fabric Management

- Fabric Mgmt Workshop (Final)
- July 8th – 9th 2003 (Sverre Jarp)
- Organized by the CERN openlab for Datagrid applications

- Aim: Understand how to create synergy between our industrial partners and LHC ...uting in the area of fabric management. The CERN talks will cover both the ...ntre (Bld. 513) and one of the LHC online farms, namely CMS.

- ...
- ...Murray, Fernando Pedone, Peter Toft
- ...rd Ferri, Kevin Gildea, Michel

- Day 2 (IT A...

- **Discussions with Intel:**
- **08:45 – 10:45 One-on-one with Intel**

- **Discussions with IBM:**
- **11:00 – 13:00 One-on-one with IBM**

- **Discussions with HP:**
- **14:00 – 16:00 One-on-one with HP**

Tomorrow

CERN
openlab for DataGrid applications

Future Events:

- **Workshop: Total Cost of Ownership**
  - **Likely date: November 2003**
  - **Possible topics:**
    - **Common vocabulary and approaches**
    - **The partners' views:**
      - **External examples**
    - **CERN's view**
      - **The P+M concept**
      - **Recent CERN acquisitions**

- **Symposium: Rational Use of Energy in Data Centres**
  - **Dates:**
    - **Monday 13 and Tuesday 14 October (during Telecom!)**
  - **Venue:**
    - **CERN IT Division, host is CERN openlab, funding from the State of Geneva (service cantonale de l'energie)**
  - **Agenda:**
    - **Conference for 60 people on 13th**
    - **Two expert workshops on 14th morning/afternoon**
    - **Results of workshop to be presented at Telecom (not confirmed)**
  - **Keywords:**
    - **benchmarking energy consumption, case study of Swisscom, research projects, low power data centers, constraints and business environment, policy and strategy**

Activities (revisited)

- ✓ **Since October 2002**
  - ✓ **Cluster installation**
  - ✓ **Cluster automation**
  - ✓ **Middleware**
  - ✓ **Compiler installations**
  - ✓ **Application porting**
  - ✓ **Benchmarking**
  - ✓ **Data Challenges**
    - ✓ **1 GB/s to tape**
    - ✓ **10 Gb/s back-to-back**
    - ✓ **10 Gb/s through ER16's**
  - ✓ **Thematic workshops**
  - ✓ **First storage subsystem investigations**
  - ✓ **A toe into Grid water with Globus**