CERN
openlab for DataGrid applications

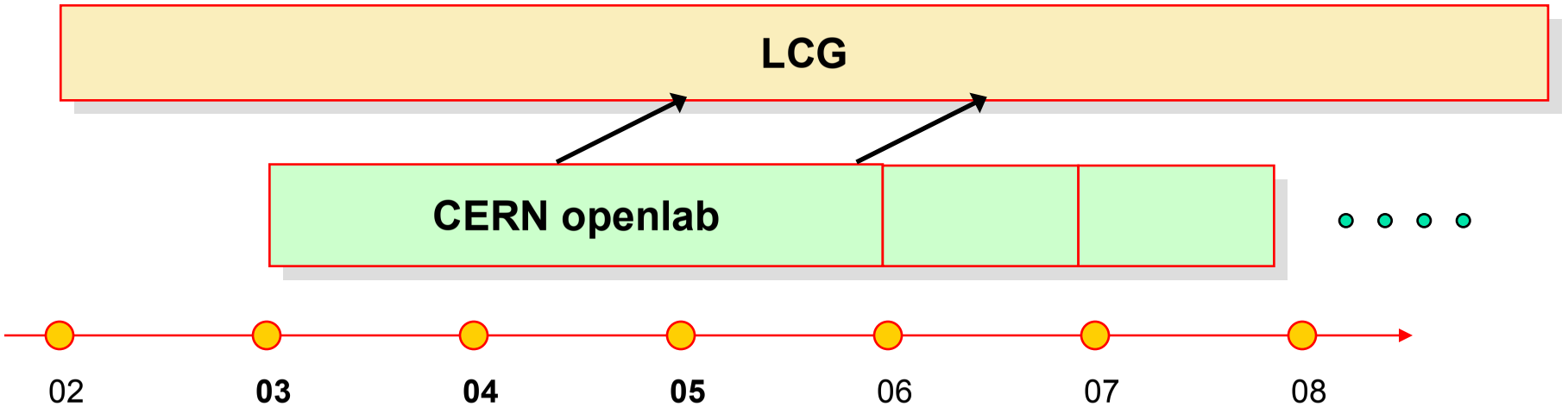
1954-2004 CERN

CERN openlab for DataGrid applications

Sverre Jarp
CERN openlab CTO
IT Department, CERN

CERN openlab

- **Department's main R&D focus**
- **Framework for collaboration with industry**
- **Evaluation, integration, validation**
  - of cutting-edge technologies
- **Initially a 3-year lifetime**
  - Later: Annual renewals



CERN
openlab for DataGrid applications

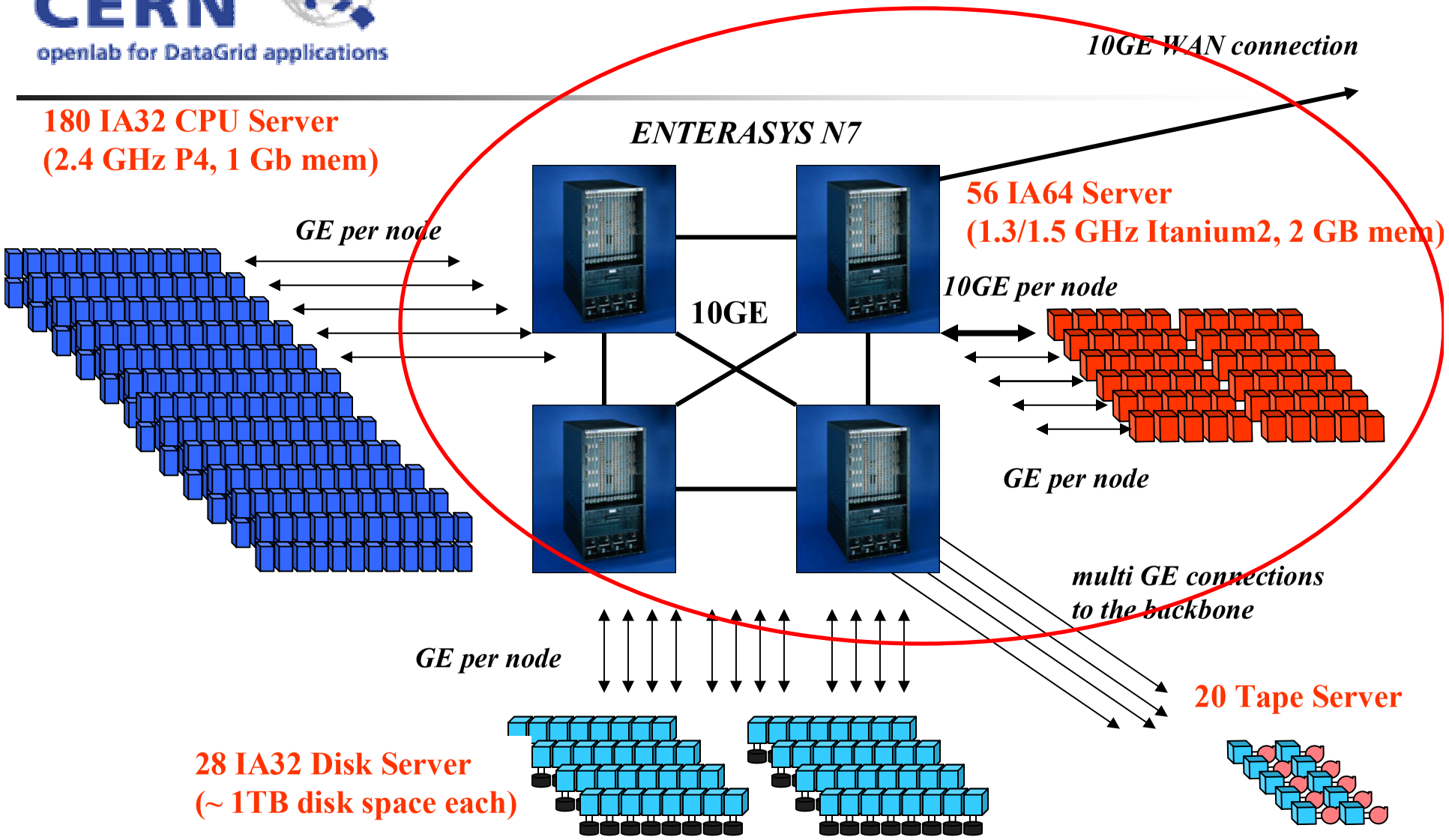
Openlab sponsors

- **5 current partners**
  - **Enterasys:**
    - **10 GbE core routers**
  - **HP:**
    - **Integrity servers (103 * 2-ways, 2 * 4-ways)**
    - **Two fellows (co-sponsored with CERN)**
  - **IBM:**
    - **Storage Tank file system (SAN FS) w/metadata servers and data servers (currently with 28 TB)**
  - **Intel:**
    - **64-bit Itanium processors & 10 Gbps NICs**
  - **Oracle:**
    - **10g Database software w/add-on's**
    - **Two fellows**
- **One contributor**
  - **Voltaire**
    - **96-way Infiniband switch**

CERN
openlab for DataGrid applications

The opencluster in its new position in the Computer Centre

Integration with the LCG testbed



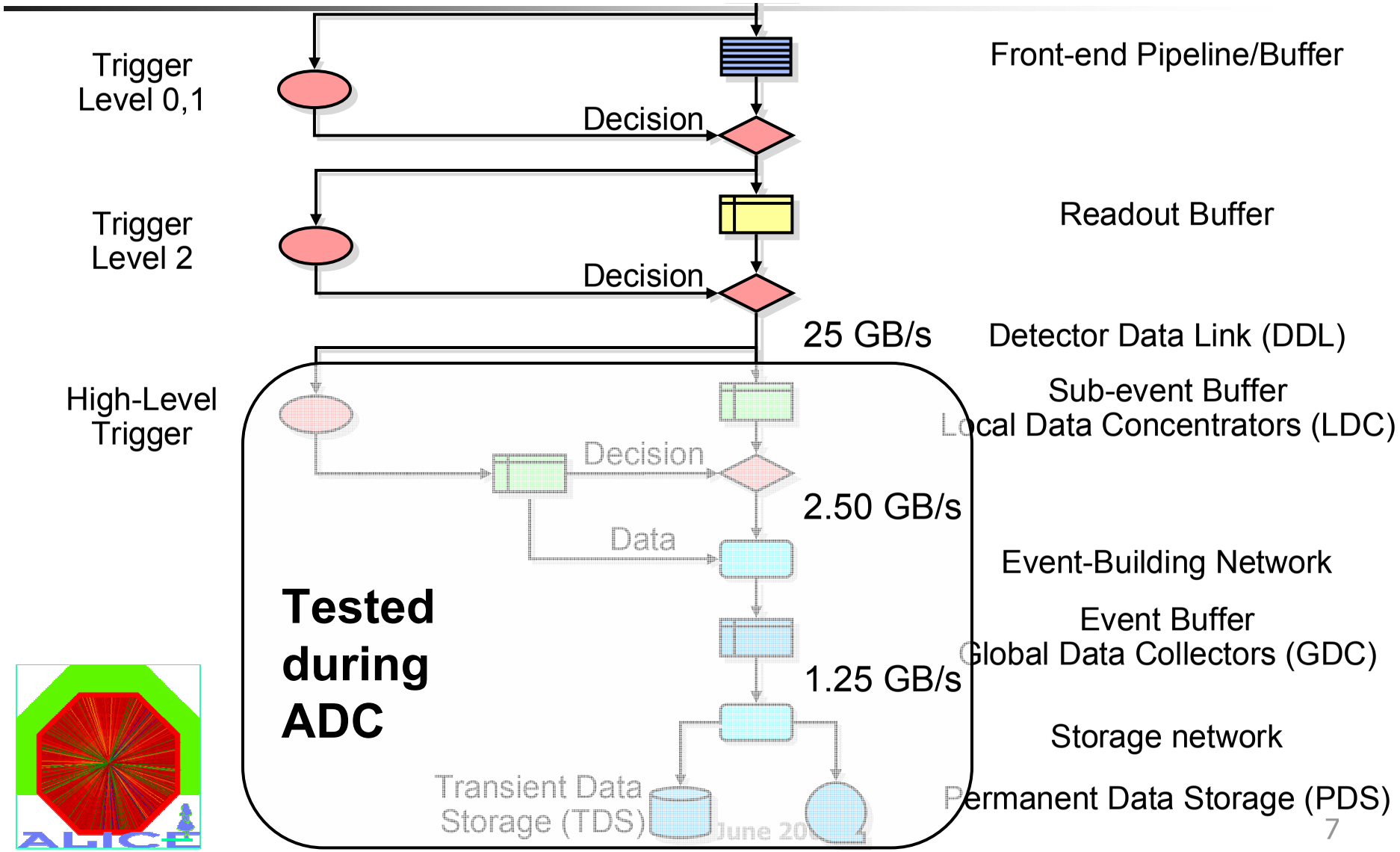
New High Througput Prototype (→ Feb. 2004)

CERN
openlab for DataGrid applications

Recent achievements
(selected amongst many others)

- **Hardware and software**
  - **Key ingredients deployed in Alice Data Challenge V**
  - **Internet2 land speed record between CERN and CalTech**
  - **Porting and verification of CERN/HEP software on 64-bit architecture**
    - **CASTOR, ROOT, CLHEP, GEANT4, ALIROOT, etc.**
  - **Parallel ROOT data analysis**
  - **Port of LCG software to Itanium**

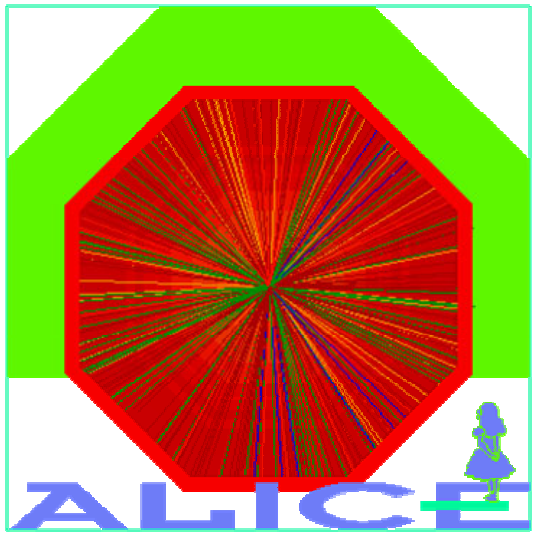
ADC V - Logical Model and Requirements



CERN
openlab for DataGrid applications

Achievements (as seen by Alice)

- Sustained bandwidth to tape:
  - Peak 350 MB/s
  - Reached production-quality level only last week of testing
  - Sustained 280 MB/s over 1 day but with interventions [goal was 300]



- ✓IA-64 from openlab successfully integrated in the ADC V

Goal for ADC VI: 450 MB/s

CERN
openlab for DataGrid applications

10 Gbps WAN tests

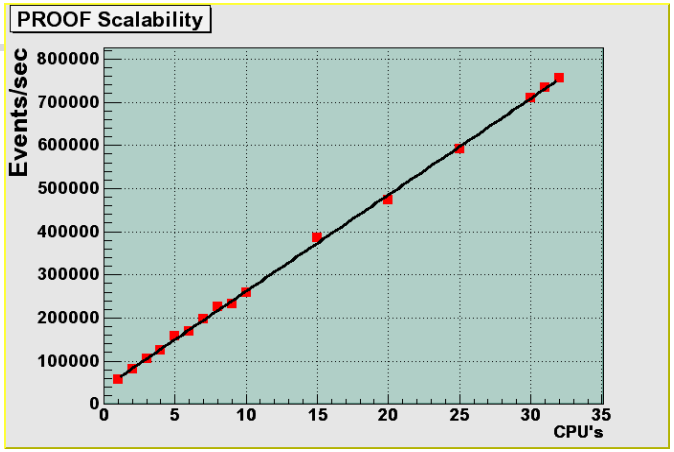
- **Initial breakthrough during Telecom-2003**
  - **with IPv4 (single/multiple) streams: 5.44 Gbps**
    - **Linux, Itanium-2 (RX 2600), Intel 10Gbps NIC**
  - **Also IPv6 (single/multiple) streams**
- **In February**
  - **Again IPv4, but multiple streams (DataTag + Microsoft): 6.25 Gbps**
    - **Windows/XP, Itanium-2 (Tiger-4), S2IO 10 Gbps NIC**
- **In June (not yet submitted)**
  - **Again IPv4, and single stream (Datatag + Openlab): 6.55 Gbps**
    - **Linux, Itanium-2 (RX2600), S2IO NIC**

openlab still has a slightly better result than NewiSys Opteron 4-way box and a heavily tuned Windows/XP

CERN
openlab for DataGrid applications

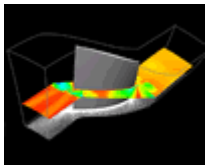
Cluster parallelization

- **Parallel ROOT Facility (PROOF):**
  - Excellent scalability with 64 processors last year
  - Tests in progress for 128 (or more) CPUs

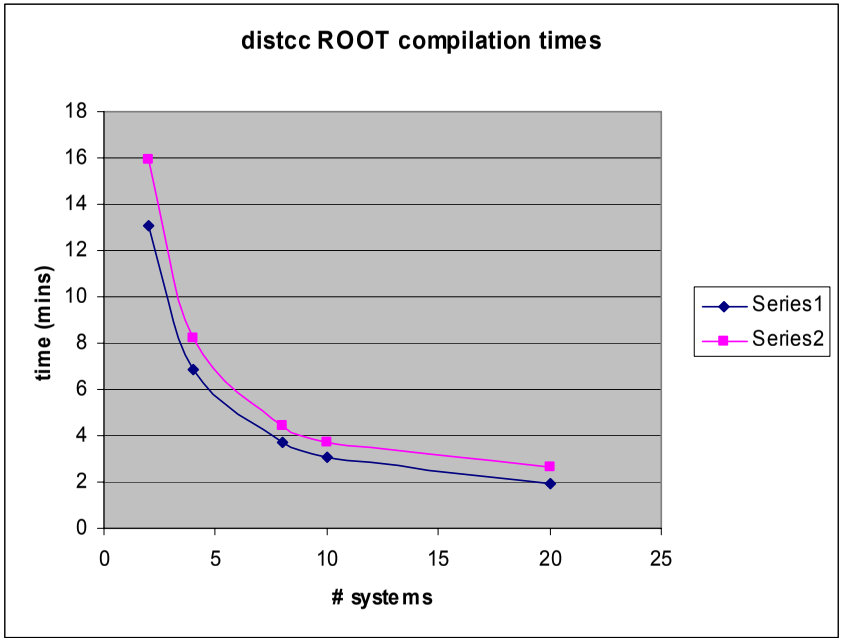


- **MPI software installed**
  - Ready for tests with ***BEAMX*** (similar program to Sixtrack)

- **Alinghi software also working**
  - Collaboration with team at EPFL
  - Uses Ansys CFX



- ***distcc*** **installed and tested**
  - Compilation time reduced both for GNU and Intel compiler



CERN
openlab for DataGrid applications

Gridification

- **A good success story:**
  - **Starting point: The software chosen for LCG (VDT + EDG) had been developed only with IA32 (and specific Red Hat versions) in mind**
    - **Consequence: Configure-files and make-files not prepared for multiple architectures. Source files not available in distributions (often not even locatable)**
  - **Stephen Eccles, Andreas Unterkircher worked for many months to complete the porting of LCG-2**
    - **Result: All major components now work on Itanium/Linux:**
      - **Worker Nodes, Compute Elements, Storage Elements, User Interface, etc.**
    - **Tested inside EIS Test Grid**
    - **Code, available via Web-site, transferred to HP sites (Initially Puerto Rico and Bristol)**
    - **Changes given back to developers**
      - **VDT now built also for Itanium systems**
  - **Porting experience summarized in white paper (on the Web)**

From now on the Grid is heterogeneous!

CERN
openlab for DataGrid applications

Storage Tank

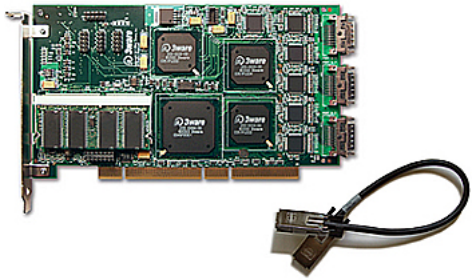
- **Random Access test (mid-March)**
  - **Scenario:**
    - **100 GB dataset, randomly accessed in ~50kB blocks**
    - **1 – 100 2 GHz P4-class clients, running 3 – 10000 "jobs"**
  - **Hardware**
    - **4 IBM x335 metadata servers**
    - **8 IBM 200i controllers, 336 SCSI disks**
    - **Added 2 IBM x345 servers as disk controllers after the test**
  - **Results**
    - **Peak data rate: 484 MB/s (with 9855 simultaneous "jobs")**
    - **After the test, special tuning, 10 servers, smaller number of clients:**
      - **705 MB/s**

Ready to be used in Alice DC VI

CERN
openlab for DataGrid applications

Next generation disk servers

- **Based on state-of-the-art equipment:**
  - **4-way Itanium server (RX4640)**
    - **Two full-speed PCI-X slots**
      - **10 GbE and/or Infiniband**



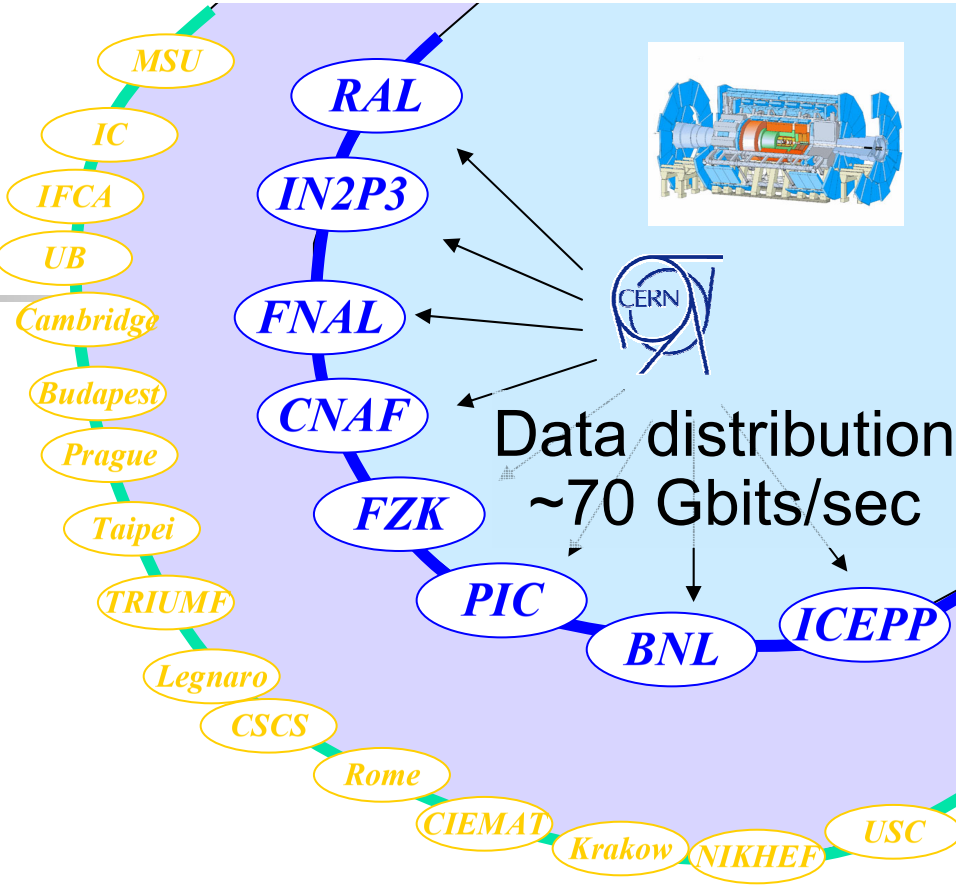
  - **Two 3ware 9500 RAID controllers**
    - **In excess of 400 MB/s RAID-5 read speed**
      - **Only 100 MB/s for write w/RAID 5**
      - **200 MB/s RAID 0**

  - **24 * S-ATA disks with 74 GB**
    - **WD740 "Raptor" @ 10k rpm**
    - **Burst speed of 100 MB/s**

Goal: Saturate 10GbE card for reading (at least 500 MB/s with standard MTU and 20 streams). Writing as fast as possible.

Data export to LCG Tier-1/-2

- **Tests (initially) between CERN and Fermilab + NIKHEF**
  - **Multiple HP Itanium servers with dual NICs**
    - **Disk to disk transfers via GridFTP**
    - **Each server:**
      - **100 MB/s IN + 100 MB/s OUT**
    - **Aggregation of multiple streams across 10 GbE link**
    - **Similar tuning as Internet2 tests**
  - **Possibly try the 4-way 10GbE server and Enterasys X-series router**



"Service Data Challenge" Stability is paramount – no longer just "raw" speed

CERN openlab for DataGrid applications

Conclusions

6 students, 4 fellows

- **CERN openlab:**
  - **Solid collaboration with our industrial partners**
  - **Encouraging results in multiple domains**
  - **We believe sponsors are getting good "ROI"**
    - **But only they can really confirm it**
  - **No risk of running short of R&D**
    - **IT Technology is still moving at an incredible pace**
  - **Vital for LCG that the "right" pieces of technology are available for deployment**
    - **Performance, cost, resilience, etc.**