HPC at CERN and the Grid

Fabrizio Gagliardi

CERN

Information Technology Division

October, 2000

F.Gagliardi@cern.ch
online system
multi-level trigger
filter out background
reduce data volume

40 MHz (40 TB/sec)
level 1 - special hardware
75 KHz (75 GB/sec)
level 2 - embedded processors
5 KHz (5 GB/sec)
level 3 - PCs
100 Hz (100 MB/sec)
data recording & offline analysis
Event Filter & Reconstruction
(figures are for one experiment)

- Data from detector - event builder
- Switch
- Computer farm
- High speed network
- Tape and disk servers
- Raw data
- Summary data

Input: 5-100 GB/sec
Capacity: 50K SI95 (~4K 1999 PCs)
Recording rate: 100 MB/sec (Alice – 1 GB/sec)

+ 1-1.25 PetaByte/year
+ 1-500 TB/year

20,000 Redwood cartridges every year (+ copy)
Offline Data and Computation for Physics Analysis

- Detector
- Raw data
- Event simulation
- Event reconstruction
- Event filter (selection & reconstruction)
- Event summary data
- Batch physics analysis
- Analysis objects (extracted by physics topic)
- Interactive physics analysis
- Processed data
Estimated CPU Capacity at CERN

Capacity that can be purchased for the value of the equipment present in 2000

technology-price curve (40% annual price improvement)

~10K SI95
1200 processors

Non-LHC

LHC

year

0 500 1,000 1,500 2,000 2,500

K SI95

Estimated DISK Capacity at CERN

Technology-price curve (40% annual price improvement)
Long Term Tape Storage Estimates

Year


TeraBytes

0 2'000 4'000 6'000 8'000 10'000 12'000 14'000

Current Experiments

COMPASS

LHC
HPC or HTC

High Throughput Computing
- mass of modest problems
- throughput rather than performance
- resilience rather than ultimate reliability

Can exploit inexpensive mass market components
But we need to marry these with inexpensive highly scalable management tools

Much in common with data mining, Internet computing facilities, ......
1960s through 1980s
- The largest scientific mainframes (Control Data, Cray, IBM, Siemens/Fujitsu)
- Time-sharing interactive services on IBM & DEC-VMS
- Scientific workstations from 1982 (Apollo) for development, final analysis

1988 -- On-line computing farms (Falcon) - joint project with Digital (microVax and Vaxstations)

1989 -- First batch services on RISC - joint project with HP (Apollo DN10.000)

1990 -- Central Simulation Facility (CSF) - 4 X mainframe capacity

1991 -- SHIFT - data intensive applications, distributed model

1993 -- First central interactive service on RISC
1994 -- 128 processor QSW (Meiko/QSW) CS2 and 72 processor IBM SP-2

1996 -- Last mainframe de-commissioned

1997 -- First batch services on PCs

1998 -- NA48 record 70 TeraBytes of data in one year
Can we scale up the current commodity-component based approach?
Computing & Storage Fabric

built up from commodity components

- Simple PCs
- Inexpensive network-attached disk
- Standard network interface
  (whatever Ethernet happens to be in 2006)

with a minimum of high(er)-end components

- LAN backbone
- WAN connection
HEP's not special, just more cost conscious

Computing & Storage Fabric
  built up from commodity components
    Simple PCs
      - Inexpensive network-attached disk
      - Standard network interface
    with a minimum of high(er)-end components
      - LAN backbone
      - WAN connection
Limited role of high end equipment

Computing & Storage Fabric
built up from commodity components
- Simple PCs
- Inexpensive network-attached disk
- Standard network interface
  (whatever Ethernet happens to be in 2006)
with a minimum of high(er)-end components

LAN backbone  WAN connection
Not everything has been commoditised yet
World Wide Collaboration

⇒ distributed computing & storage capacity
Regional Computing Centres

- Exploit established computing expertise & infrastructure
  - In national labs, universities

- Reduce dependence on links to CERN
  - full ESD available nearby - through a fat, fast, reliable network link

- Tap funding sources not otherwise available to HEP
Regional Centres - a Multi-Tier Model

CERN - Tier 0

622 Mbps

FNAL
155 Mbps

RAL
2.5 Gbps

IN2P3

Tier 1

Tier 2

Lab a

Uni a

Lab c

Uni n

MONARC report: http://home.cern.ch/~barone/monarc/RCArchitecture.html
More realistically - a Grid Topology
Summary - the basic problem

- **Scalability**
  - Thousands of processors, thousands of disks, PetaBytes of data, Terabits/second of I/O bandwidth, ....

- **Wide-area distribution**
  - WANs are and will be 1% of LANs
  - Distribute, replicate, cache, synchronise the data
  - Multiple ownership, policies, ....
  - Integration of this amorphous collection of Regional Centres
  - With some attempt at optimisation

- **Adaptability**
  - We shall only know how analysis is done once the data arrives
Are Grids a solution?

Change of orientation of US Meta-computing activity
- From inter-connected super-computers
  ... towards a more general concept of a computational Grid (The Grid – Ian Foster, Carl Kesselman)

Has initiated a flurry of activity in HEP
- US – Particle Physics Data Grid (PPDG)
- Grid technology evaluation project in INFN
- UK proposal for funding for a prototype grid
- GriPhyN – data grid proposal just approved by NSF
- NASA Information Processing Grid

DataGrid initiative launched
The GRID metaphor

- Unlimited ubiquitous distributed computing
- Transparent access to multipetabyte distributed data bases
- Easy to plug in
- Hidden complexity of the infrastructure
- Analogy with the electrical power GRID
The Grid from a Services View

Applications

- Chemistry
- Cosmology
- Environment
- Biology
- High Energy Physics

Application Toolkits

- Distributed Computing Toolkit
- Data-Intensive Applications Toolkit
- Collaborative Applications Toolkit
- Remote Visualization Applications Toolkit
- Problem Solving Applications Toolkit
- Remote Instrumentation Applications Toolkit

Grid Services (Middleware)

Resource-independent and application-independent services
- authentication, authorization, resource location, resource allocation, events, accounting,
- remote data access, information, policy, fault detection

Grid Fabric (Resources)

Resource-specific implementations of basic services
- E.g., Transport protocols, name servers, differentiated services, CPU schedulers, public key
- infrastructure, site accounting, directory service, OS bypass
Five Emerging Models of Networked Computing From

-The Grid-

- Distributed Computing
  - // synchronous processing
- High-Throughput Computing
  - // asynchronous processing
- On-Demand Computing
  - // dynamic resources
- Data-Intensive Computing
  - // databases
- Collaborative Computing
  - // scientists

Local fabric

- **Management of giant computing fabrics**
  - auto-installation, configuration management, resilience, self-healing

- **Mass storage management**
  - multi-PetaByte data storage, “real-time” data recording requirement, active tape layer - 1,000s of users

Wide-area - building on an existing framework & RN (e.g. Globus, Geant and high performance network R&D)

- **Workload management**
  - no central status
  - local access policies

- **Data management**
  - caching, replication, synchronisation
  - object database model

- **Application monitoring**
HEP Data Grid Initiative

- European level coordination of national initiatives & projects
- Principal goals:
  - Middleware for fabric & Grid management
  - Large scale testbed - major fraction of one LHC experiment
  - Production quality HEP demonstrations
    - "mock data", simulation analysis, current experiments
  - Other science demonstrations
- Three year phased developments & demos
- Complementary to other GRID projects
  - EuroGrid: Uniform access to parallel supercomputing resources
- Synergy being developed (GRID Forum, Industry and Research Forum)
Participants

- Main partners: CERN, INFN(I), CNRS(F), PPARC(UK), NIKHEF(NL), ESA-Earth Observation
- Other sciences: KNMI(NL), Biology, Medicine
- Industrial participation: CS SI/F, DataMat/I, IBM/UK
- Associated partners: Czech Republic, Finland, Germany, Hungary, Spain, Sweden (mostly computer scientists)
- Formal collaboration with USA established
- Industry and Research Project Forum with representatives from:
  - Denmark, Greece, Israel, Japan, Norway, Poland, Portugal, Russia, Switzerland
Status

- Prototype work already started at CERN and in most of collaborating institutes (Globus initial installation and tests)

- Proposal to the EU positively reviewed at the end of July, 9.8 M Euros (covering 1/3 of total investment), 3 years contract being negotiated now

- Expect start of the project, January next year
Conclusions

- The Grid is a useful metaphor to describe an appropriate computing model for LHC and future HEP computing
- Middleware, APIs and interface general enough to accommodate many different models for science, industry and commerce
- Still important R&D to be done
- Perfect field for multidisciplinary collaboration (computer science, physics and other sciences)
- If successful could develop next generation Internet computing
- Major funding agencies prepared to fund large testbeds in USA, EU and Japan
- Excellent opportunity for HEP computing to deploy a sustainable HPC model