

MC Production in Canada

Pierre Savard

University of Toronto and TRIUMF

IFC Meeting

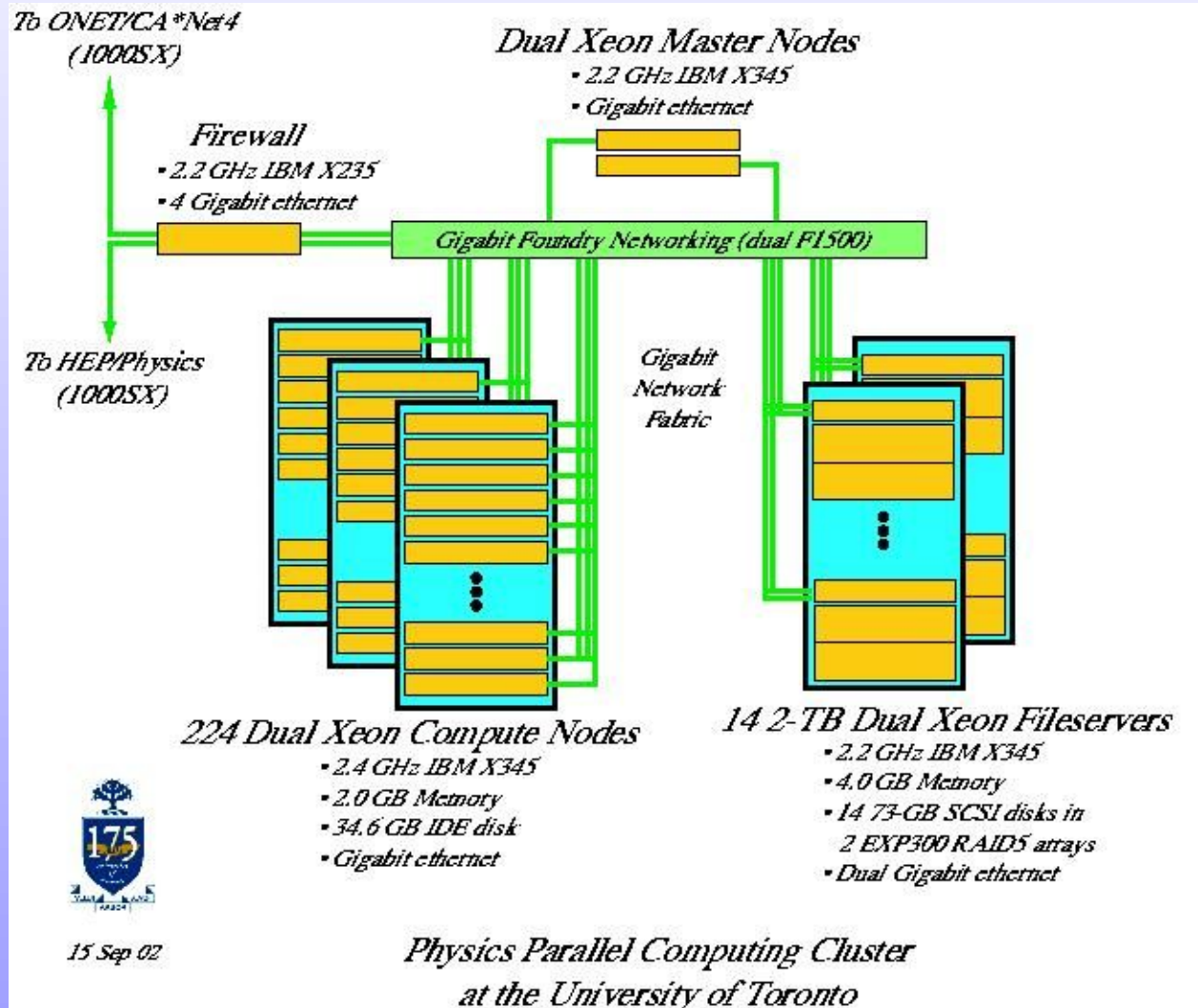
October 2003

Overview

- Collaboration has need for production of large, common Monte Carlo simulation datasets
 - Frank Wuerthwein originally setup MC production group to organize production of “official” MC datasets
 - Production of datasets were done in large part on CDF’s reconstruction and analysis clusters at FNAL
- CDF Canadian institutions have large beowulf clusters with excellent connectivity to FNAL
- New investigators (Pinfold, Savard, Warburton) proposed to exploit these resources to produce most of CDF’s official MC
- Granted operation funds by Canadian funding agency, signed MOU with CDF management

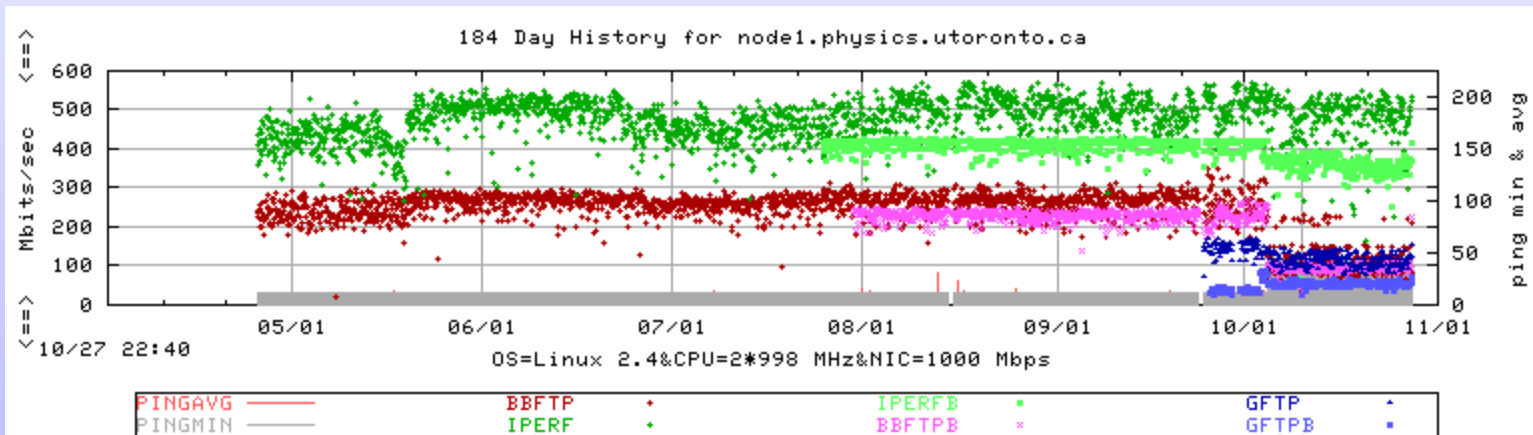
Resources in Canada

Toronto: last year acquired large linux cluster with ~450 cpus



Resources in Canada

- Toronto connectivity:
 - Directly connected to Canada research network (Gigabit across country and to STARLIGHT)



- Alberta: Thor multi-processor facility
 - 170 nodes of dual 2GHz cpus
 - 4 TB of disk
 - Network fabric is GigE
- McGill: machine comparable to Toronto cluster, has not been used yet

MC Production MOU

- Dedicate a MINIMUM of 250 GHz of equivalent cpus to produce official CDF MC datasets for 2 years
- Assume coordination of MC production
- Assume responsibility for transfer of data to FNAL and to tape (all official MC datasets go in DFC)
- Dedicate necessary human resources to coordinate and manage production at computing facilities (1.7 FTE)

MC Production in Canada

- **Proposal a good match to our resources: MC requires a lot of cpu (generation + simulation + reconstruction), good bandwidth (to push data back)**
- **Toronto farm setup as a CDF FNAL machine: Fermi Linux, Fermi batch system. Main other user of Toronto and Alberta clusters is ATLAS for data/GRID challenges**
- **Note that human resource component should not be underestimated: need to prepare and submit jobs, to do data handling, to keep track of logs, transfer data etc. This does not include machine admin or maintenance**

Delivered MC Production

- Last year, produced more than half of official MC for CDF. **Just for Summer conferences:**
 - produced 43 of 55 million events
 - transferred ~10 TB at 15-20 Mbytes/s (average rate)
 - 260 datasets written to tape
- We operated beyond the proposed “minimum”:
 - During period of high demand, typically allocated more than 300 cpus in Toronto and 50 cpus in Alberta
- Also regularly produce large samples to test new simulation releases.
- **Total of ~80 million events produced last year**

Conclusions

- Large-scale Monte Carlo dataset production at remote institutions has been a success
- Reduces load on FNAL reconstruction and analysis clusters
- Have followed a detector hardware model:
 - provide hardware component and necessary human resources to operate it.
 - Computers are funded by Canadian government which also provides operation funds