Networking Issues for CDF/Canada Computing

Pekka K. Sinervo

Department of Physics

University of Toronto

15 Apr 2003

May 02

CDF Computing Problem

- **Tevatron collecting data to 2006**
  - **Maxium data logging rate 80 Hz**
  - **With running efficiency of 30%**
    - ✦ Collect 0.8B events/year
    - ✦ Results in 2.0B events by 2004
      - **Integrated luminosity of 2.5 fb^{-1}**
  - **Plan on 0.6B MC events by 2004**

- **With 30% reprocessing, requires**
  - **560 GHz processors for data farms**
  - **200 GHz processors for MC farms**

- **Creates a flood of physics data**
  - **100 TB/fb^{-1} of reconstructed data**
 - **1.1 PByte archival tape storage**

CDF/Canada Networking

- **Two sets of local requirements**

  1) **Computing for local physics analysis, calibration and simulation**
     - Difficult to assess, but experience shows that one needs:
       - **Linux w/s for each physicist**
       - **Access to fast turn-around batch**
       - **Need of order 100 GB/activity**
     - Many analyses limited by disk read speed
       - **100BaseT network inadequate**

  2) **CDF MC Production Facility**
     - Planned to produce 150M events/year
     - Store MC DSTs at FNAL
       - **At 0.15 MB/event, need approximately 1 MB/s DC**
       - **Data analysis needs add an additional 0.5 MB/s DC**

Networking Layout

■ **Components of the network**

  – **Work stations and disk (fish cluster)**

    ✦ 100BaseT providing 5-7 MB/s

    ✦ NFS-mounted disk on disk/compute servers

  – **Big Mac Cluster**

    ✦ Internally GigE networking

    ✦ Connection through firewall

      – **1000SX connectivity**
      – **Data transfer >70 MB/s**

  – **Wide-Area Network**

    ✦ Connect to CA*Net4 Gigapop

    ✦ CA*Net4 to StarLight (Chicago)

    ✦ ESNet to Fermilab (via StarTap)

      – **Limitation is ESNET**
      – **OC3 (156Mbps) + OC12 (622 Mbps) link from StarTap**
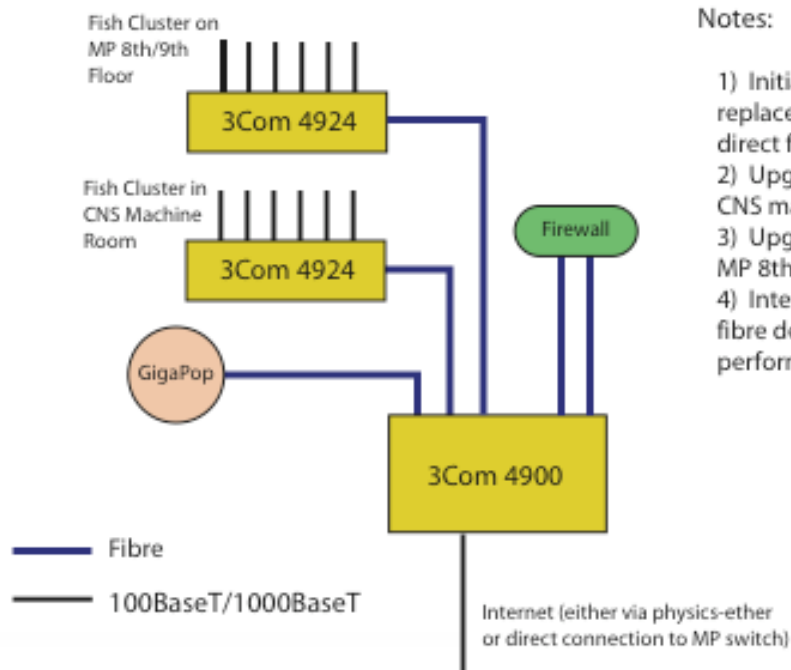
Toronto Fish Cluster

- **Have small Linux cluster**
  - Worker nodes (8) and w/s (12)
  - Disk space of 2.0TB
  - Connected via 100BaseT
  - One server (piranha) has 1000SX GBIC interface

University of Toronto Physics   5
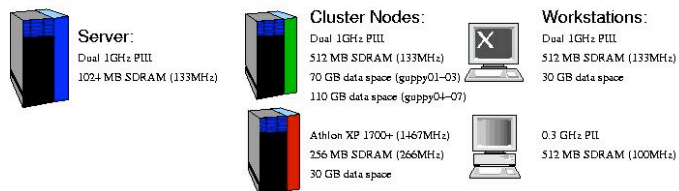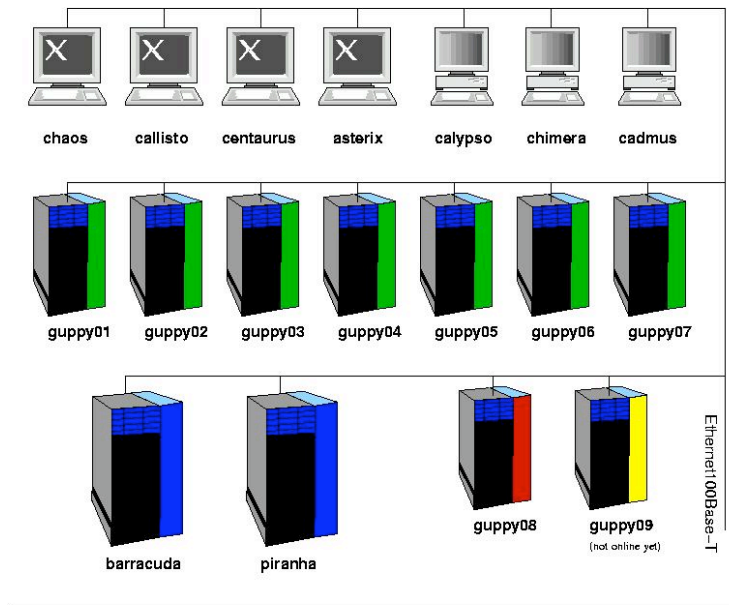
Local Network Plan

- **Deployed 3Com 4900 switch**
  - **Provides connectivity between Big Mac, GigaPop, and HEP cluster**
  - **Additional 3Com 4924 switches to be added to get beyond 100BaseT**
    - One for servers on 3rd floor
    - One for 8th/9th floor w/s

FINAL CONFIGURATION

Fish Cluster on MP 8th/9th Floor → 3Com 4924

Fish Cluster in CNS Machine Room → 3Com 4924

GigaPop

Firewall

3Com 4900

Notes:

1) Initial configuration would replace 4924 switches with direct fibre connection to Piranha.
2) Upgrade path #1: Add 4924 to CNS machine room.
3) Upgrade path #2: Add 4924 to MP 8th floor.
4) Internet connection could be fibre depending on cost and expec performance beyond MP switch.

— Fibre

— 100BaseT/1000BaseT

Internet (either via physics-ether or direct connection to MP switch)

External Connectivity

- **UofT Connectivity is complex**
  - **Actual I/O from desktop to Fermilab was limited to 1-2 MB/s**
  - **Limitation is artificial**
    - Gateway to CA*Net4 is 15 Mbps
  - **Now have dedicated connection to GigaPop**
    - In principle have >2 Gbps service in place
    - Have been monitoring bandwidth since Nov 2002
    - Readily get 20 MB/s disk-to-disk
      - **3.5 GB transfer test takes 3 min**
    - Monitoring using IEPM software
      - **Shows consistent and clear pipe to Fermilab**
      - **Limitation remains ESNET**

IEPM Monitoring

- **Software installed on piranha**
  - Internet End-to-End Performance Monitoring package developed at SLAC
  - Monitors status every 15 min

60 Day History for node1.physics.utoronto.ca
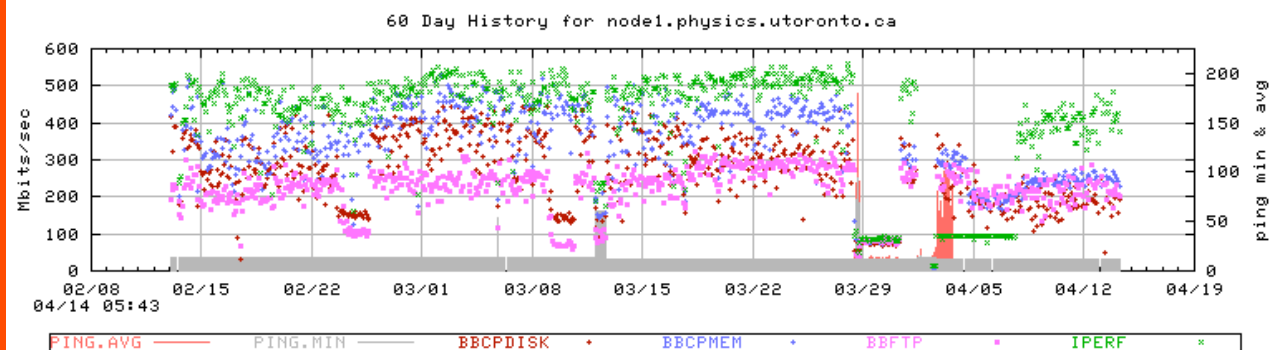
  - **Comments:**
    - ✦ Bandwidth is quite stable
      - Differences in actual measures are significant
    - ✦ Actual throughput measured via ftp is closer to 160 Mbps
    - ✦ Can see effect of 3Com 4900
      - Note well tuned at this point

Next NetworkingSteps for CDF/Canada

- **Start monitoring Alberta & McGill links**
    - **To Toronto and Fermilab**
    - **Goal is to have clear pipes between all four institutions**

- **Develop most robust and efficient data transfer tools**
    - **ftp is clearly limited**
        - Session limited to 2.5 MB/s
        - But can do multiple sessions
    - **bbcp protocol looks reasonable**
        - Gets 25 MB/s with single session

- **Locally move to 1000BaseT network adiabatically as needed**