DØ Data Handling Operational Experience

Roadmap of Talk

- DØ overview
- Computing Architecture
- Operational Statistics
- Operations strategy and experience
- Challenges and Future Plans
- Summary

CHEP03
UCSD
March 24-28, 2003
Lee Lueking
The DØ Experiment

- D0 Collaboration
  - 18 Countries; 76 institutions
  - 600 Physicists
- Detector Data (Run 2a end mid ‘04)
  - 1,000,000 Channels
  - Event size 250KB
  - Event rate 25 Hz avg.
  - Est. 2 year data totals (incl Processing and analysis): 1 x 10^9 events, ~1.2 PB
- Monte Carlo Data (Run 2a)
  - 6 remote processing centers
  - Estimate ~0.3 PB.
- Run 2b, starting 2005: >1PB/year
DØ Data Handling is a Successful Worldwide Effort

Thanks to the efforts of many people

- The SAM team at FNAL: Andrew Baranovski, Diana Bonham, Lauri Carpenter, Lee Lueking, Wyatt Merritt*, Carmenita Moore, Igor Terekhov, Julie Trumbo, Sinisa Veseli, Matthew Vranicar, Stephen P. White, Kin Yip (BNL). (*project co-lead with Jeff Tseng from CDF)
- Major contributions from –Amber Boehnlein (D0 Offline Computing Leader), Iain Bertram (Lancaster), Chip Brock (MSU), Jianming Qian (UM), Rod Walker (IC), Vicky White (FNAL)
- CD Computing and Communication Fabric Dept. (CCF), in particular the Enstore Team, and Farms Support Group
- CD Core Computing Support (CCS) Database Support Group (DSG) and Computing and Engineering for Physics Applications (CEPA) Database Applications Group (DBS) for database support
- CD D0 Department, D0 Operations Team at Fermilab
- CAB and CLueD0 administrators and support teams
- Sam Station Administrators, and SAM Shifters Worldwide
- MC production teams: Lancaster UK, Imperial College UK, Prague CZ, U Texas Arlington, Lyon FR, Amsterdam NL,
- GridKa Regional Analysis Center at Karlsruhe, Germany: Daniel Wicke, Christian Schmitt, Christian Zeitnitz
SAM Data Management System

- SAM is Sequential data Access via Meta-data
- Est. 1997
- Flexible and scalable distributed model
- Field hardened code
- Reliable and Fault Tolerant
- Adapters for many batch systems: LSF, PBS, Condor, FBS
- Adapters for mass storage systems: Enstore, (HPSS, and others planned)
- Adapters for Transfer Protocols: cp, rcp, scp, encp, bbftp, GridFTP.
- Useful in many cluster computing environments: SMP w. compute servers, Desktop, private network (PN), NFS shared disk,…
- Ubiquitous for DØ users

SAM Station – 1. Collection of SAM servers which manage data delivery for a node or cluster
2. The node or cluster hardware itself
Overview of DØ Data Handling

Integrated Files Consumed vs Month (DØ)

4.0 M Files Consumed

Integrated GB Consumed vs Month (DØ)

1.2 PB Consumed

Summary of DØ Data Handling

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered Users</td>
<td>600</td>
</tr>
<tr>
<td>Number of SAM Stations</td>
<td>56</td>
</tr>
<tr>
<td>Registered Nodes</td>
<td>900</td>
</tr>
<tr>
<td>Total Disk Cache</td>
<td>40 TB</td>
</tr>
<tr>
<td>Number Files - physical</td>
<td>1.2M</td>
</tr>
<tr>
<td>Number Files - virtual</td>
<td>0.5M</td>
</tr>
<tr>
<td>Robotic Tape Storage</td>
<td>305 TB</td>
</tr>
</tbody>
</table>

Mar2002

Mar2003

Regional Center

Analysis site

d0db.fnal.gov/sam

CHEP 03 UCSD
Data In and out of Enstore
(robotic tape storage) Daily Feb 14 to Mar 15

1.3 TB incoming
2.5 TB outgoing

Enstore Talk, Cat. 3 Tuesday
## DØ SAM Station Summary

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Nodes/cpu</th>
<th>Cache</th>
<th>Use/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central-analysis</td>
<td>FNAL</td>
<td>128 SMP*, SGI Origin 2000</td>
<td>14 TB</td>
<td>Analysis &amp; DØ code development</td>
</tr>
<tr>
<td>CAB (CA Backend)</td>
<td>FNAL</td>
<td>16 dual 1 GHz + 160 dual 1.8 GHz</td>
<td>6.2 TB</td>
<td>Analysis and general purpose</td>
</tr>
<tr>
<td>FNAL-Farm</td>
<td>FNAL</td>
<td>100 dual 0.5-1.0 GHz + 240 dual 1.8 GHz</td>
<td>3.2 TB</td>
<td>Reconstruction</td>
</tr>
<tr>
<td>CLueD0</td>
<td>FNAL</td>
<td>50 mixed PIII, AMD. (may grow &gt;200)</td>
<td>2 TB</td>
<td>User desktop, General analysis</td>
</tr>
<tr>
<td>DØkarlsruhe (GridKa)</td>
<td>Karlsruhe, Germany</td>
<td>1 dual 1.3 GHz gateway, &gt;160 dual PIII &amp; Xeon</td>
<td>3 TB</td>
<td>General/Workers on PN. Shared facility</td>
</tr>
<tr>
<td>Nijmegen</td>
<td>Nijmegen, Netherlands</td>
<td>1 dual 1.8 GHz gateway, 6 x dual 930MHz</td>
<td>1 TB</td>
<td>Analysis/ workers on PN</td>
</tr>
<tr>
<td>Many Others &gt; 4 dozen</td>
<td>Worldwide</td>
<td>Mostly dual PIII, Xeon, and AMD XP</td>
<td></td>
<td>MC production, gen. analysis, testing</td>
</tr>
</tbody>
</table>

*IRIX, all others are Linux
Station Stats: GB Consumed
Daily Feb 14 – Mar 15

Central-Analysis
2.5 TB
Feb 22

ClueD0
270 GB
Feb 17

FNAL-farm
1.1 TB
Mar 6

CAB
>1.6 TB
Feb 28
Station Stats: MB Delivered/Sent
Daily Feb 14 – March 15

Central-Analysis
1 TB
Feb 22
Delivered
2.5 TB
Feb 22
Consumed

FNAL-farm
1.2 TB
Mar 6

ClueD0
150 GB
Feb 17

CAB
600 GB
Feb 28

1.1 TB
Mar 6
Consumed

1.6 TB
Feb 28
Consumed
FNAL-farm Station and CAB CPU Utilization
Feb 14 – March 15

- 600 CPUs
- 50% Utilization
- CAB Usage will increase dramatically in the coming months

Also, see CLued0 Talk in Section 3, Monday
Data to and from Remote Sites

Station Configuration
• Replica location
  • Prefer
  • Avoid
• Forwarding
  • File stores can be forwarded through other stations
• Routing
  • Routes for file transfers are configurable

Extra-domain transfers use bbftp or GridFTP (parallel transfer protocols)
DØ Karlsruhe Station at GridKa

Monthly Thumbnail Data Moved to GridKa

1.2 TB in Nov 2002

Cumulative Thumbnail Data Moved to GridKa

5.5 TB since June 2002

The GridKa SAM Station uses shared cache config. with workers on a private network

This is our first Regional Analysis Center (RAC). See DØ RAC Concept talk, Category 1, Tuesday.

Resource Overview:
- Compute: 95 x dual PIII 1.2GHz, 68 x dual Xeon 2.2 GHz. DØ requested 6%. (updates in April)
- Storage: DØ has 5.2 TB cache. Use of ~100TB MSS. (updates in April)
- Network: 100Mb connection available to users.
- Configuration: SAM w/ shared disk cache, private network, firewall restrictions, OpenPBS, Redhat 7.2, k2.4.18, DØ software installed.
SAM Shift Stats

Overview

- In operation since summer 2001
- Kin Yip (BNL) is current shift coordinator
- 27 general shifters from 5 time zones.
- 7 expert shifters at FNAL
- Experts still carry much of the load.
- Problems range from simple user questions to installation issues, hardware, network, bugs...

Weekly SAM Problem Resolution

Shifter Time Zones

SAM Problem Resolution

In operation since summer 2001
Kin Yip (BNL) is current shift coordinator
27 general shifters from 5 time zones.
7 expert shifters at FNAL
Experts still carry much of the load.
Problems range from simple user questions to installation issues, hardware, network, bugs...
## SAM At A Glance
### D0 Production Environment
This page generated on 16 Mar 2003 13:29:47

**Other Sam Diagnostics**

### SAM Stations:

#### Monitor Level: Critical

<table>
<thead>
<tr>
<th>Host</th>
<th>Version</th>
<th>Up Since</th>
</tr>
</thead>
<tbody>
<tr>
<td>d0clb-gb-1.fnal.gov</td>
<td>v4_2_1_12</td>
<td>07 Mar 2003 17:13:43</td>
</tr>
<tr>
<td>d0clr-gb-1.fnal.gov</td>
<td>v4_2_1_12</td>
<td>07 Mar 2003 17:13:09</td>
</tr>
</tbody>
</table>

#### Monitor Level: High

<table>
<thead>
<tr>
<th>Host</th>
<th>Version</th>
<th>Up Since</th>
</tr>
</thead>
<tbody>
<tr>
<td>d0mino.fnal.gov</td>
<td>v4_2_1_24</td>
<td>14 Mar 2003 16:47:29</td>
</tr>
<tr>
<td>d0bbin.fnal.gov</td>
<td>v4_2_1_31</td>
<td>05 Mar 2003 11:54:19</td>
</tr>
</tbody>
</table>

#### Monitor Level: Normal

<table>
<thead>
<tr>
<th>Host</th>
<th>Version</th>
<th>Up Since</th>
</tr>
</thead>
<tbody>
<tr>
<td>d0mino-sam.fnal.gov</td>
<td>v4_2_1_40</td>
<td>14 Mar 2003 13:48:19</td>
</tr>
<tr>
<td>no smaster/ffs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ccd4c.in2p3.fr</td>
<td>v4_2_1_24</td>
<td>10 Mar 2003 09:53:54</td>
</tr>
<tr>
<td>ccd4c.in2p3.fr</td>
<td>v4_2_1_24</td>
<td>10 Mar 2003 09:53:57</td>
</tr>
<tr>
<td>d0mino.fnal.gov</td>
<td>v4_2_1_39</td>
<td>04 Mar 2003 23:04:46</td>
</tr>
<tr>
<td>higgs.fis.cernet-station</td>
<td>v4_2_1_31</td>
<td>24 Feb 2003 11:59:47</td>
</tr>
<tr>
<td>flotsam-clued0.fnal.gov</td>
<td>v4_2_1_40</td>
<td>07 Mar 2003 11:43:26</td>
</tr>
<tr>
<td>d0oral1.fnal.gov</td>
<td>v4_2_1_20</td>
<td>25 Feb 2003 08:30:34</td>
</tr>
<tr>
<td>no smaster/ffs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

[Image of a webpage showing SAM At A Glance with details on SAM stations and their statuses.]
Challenges

• Getting SAM to meet the needs of DØ in the many configurations is and has been an enormous challenge. Some examples include…
  – File corruption issues. Solved with CRC.
  – Preemptive distributed caching is prone to race conditions and log jams. These have been solved.
  – Private networks sometimes require “border” naming services. This is understood.
  – Additional simplicity and generality are provided in the NFS shared cache configuration, at the price of scalability (star configuration). This works.
  – Installation procedures for the station servers have been quite complex. They are improving and we plan to soon have “push button” and even “opportunistic deployment” installs.
  – Lots of details with opening ports on firewalls, OS configurations, registration of new hardware, and so on.
  – Username clashing issues. Moving to GSI and Grid Certificates.
  – Interoperability with many MSS.
  – Network attached files. Sometimes, the file does not need to move to the user.
Stay Tuned for SAM-Grid
The best is yet to come…

JIM Talk
In Cat. 1
Igor Terekhov
Summary

- The DØ Data Handling operation is a complex system involving a worldwide network of infrastructure and support.
- SAM provides flexible data management solutions for many hardware configurations, including clusters in private networks, shared NFS cache, and distributed cache. It also provides configurable data routing throughout the install base.
- The software is stable and provides reliable data delivery and management to production systems at FNAL and worldwide. Many challenging problems have been overcome to achieve this goal.
- Support is provided through a small group of experts at FNAL, and a network of shifters throughout the world. Many tools are provided to monitor the system, detect and diagnose problems.
- The system is continually being improved, and additional features are planned as the system moves beyond data handling to complete Grid functionality in the SAM-Grid project (a.k.a. SAM + JIM).