Reprocessing DØ data with SAMGrid

All Hands Meeting, 19-22 Sept. 2005 in Nottingham

Frédéric Villeneuve-Séguier, IMPERIAL COLLEGE London
On behalf of the DØ Collaboration and the SAMGrid team
Outline

- The DØ experiment at Fermilab
  - The Tevatron collider and the DØ collaboration
  - Physics programme and challenges
  - Computing model and application
  - Data flow and volume

- SAMGrid architecture
  - SAM (data handling mechanism)
  - SAMGrid infrastructure and services

- Reprocessing DØ data using SAMGrid
  - Data reprocessing task
  - Application flow and site certification
  - Monitoring, status and problems encountered
  - UK Regional centers
  - SAMGrid interoperability

- Conclusion and prospects
The DØ experiment at Tevatron

- Fermilab (FNAL) : research in High Energy Physics and related disciplines (~ 50 miles from Chicago, USA).
- ~30 experiments (proton-antiproton, fixed targets, neutrino physics, ...)
- Tevatron, proton-antiproton collider, highest energy collider currently operational.
- Collisions studied at 2 interaction points (CDF, DØ).
- New data taking period started in march 2001 ("Run II").

- DØ is a worldwide collaboration of about 650 physicists from over 20 participating countries.
- "Run I" (1992-1996) lead to “top” quark discovery.
- 132 physics papers have been published up to now ( >30 with “Run II” data)
- The DØ detector has been significantly upgraded to improve its performances for “Run II”

F. Villeneuve-Seguier @ AHM 2005
The DØ detector has been designed to study a large spectrum of High Energy Physics phenomena:

- Higgs searches
- CP violation (heavy flavour physics)
- QCD studies (high transverse momentum jets)
- Electroweak precision measurements
- Physics beyond the Standard Model (SUSY, lepto-quarks, extra-dimensions, ...)

The “Run II” proton-antiproton environment is much busier than “Run I”.

- ~ 10 million proton-antiproton collision per second
- The complexity of such an environment is a real challenge at several levels:
  - Hardware (calibration, measurement, trigger, ...)
  - Physics objects reconstruction
  - Computing resources (storage, computing power, management, ...)

Reprocessing DØ data with SAMGrid – AHM 2005
Computing model and data/MC production

- From distributed computing ...
  - Originally, distributed computing at “remote” sites was the solution adopted to enable access to large amount of data and Monte-Carlo simulated data.
  - All DØ Monte-Carlo production was done outside of FNAL, using resources shared with other experiments.
  - Since the beginning of “Run II”, on-demand data distribution for remote analysis is done through the use of a data grid (SAM).
  - Build activity around Regional centers.

- ... to computational grid
  - The DØ computing model has progressively evolved from distributed computing to a computational grid model (SAMGrid), based on the use of common standard tools on the grid.
  - SAMGrid has been the default for both Monte-Carlo production and data reprocessing since winter 2003/2004.
  - Next step: Other production tasks and user analysis (in order of increasing complexity).
Data flow and volume

- Proton-antiproton collisions producing potentially interesting events are selected through the trigger system.
- Digitized output from each sub-detector is then stored in the “so called” RAW data format.
- RAW data files are then processed through the DØ reconstruction software to create higher level physics objects (tracks, electrons, muons, ...).
- Initially, the output from the previous stage was stored in a DST format, but in order to reduce the storage volume, reconstructed events are now condensed in the “thumbnail” format (TMB).
- In addition, metadata for both RAW and TMB files are created.
- Data files and metadata are then stored on tape.

Since the beginning of “Run II”, more than 1000 million collisions have been recorded.
- This represents about 500 TB of data.
- Large amount of Monte-Carlo simulated data are also required to perform physics analysis.
- Some 75 million events of MC data were generated this year.
SAMGrid architecture

**JIM**
- Job handling and submission

**SAM**
- Job monitoring and information

**SAM Grid Architecture**

JIM + SAM = SAMGRID
SAM : data handling management

- SAM (Sequential Access to Metadata) originally developed by DØ and the Fermilab Computing Division (now used by other experiments as well).

- Data handling mechanism, organized as a set of servers (SAM stations).

- SAM allows:
  - storage
  - Delivery utilities on request
  - Methods of job submission for local or grid-aware systems

- SAM provides a data grid to the SAMGrid architecture
Reprocessing DØ data with SAMGrid – AHM 2005

SAM plots

- ~ 50 SAM sites worldwide
- Over 2.5 PB (~ 50B events) consumed last year
- Up to 300 TB moved per month

samTV - Sam Snapshot Summaries

SAM TV – monitor SAM and SAM stations

http://www-clued0.fnal.gov/%7Esam/samTV/current/

F. Villeneuve-Seguier @ AHM 2005
Reprocessing DØ data with SAMGrid – AHM 2005

SAMGrid services and deployment

- 3 components of SAMGrid:
  - Data handling
  - Job monitoring and information
  - Job handling

(SAM + JIM = SAMGrid)

SAMGrid had to develop its own “job manager” (batch adapter) to have a proper interface between “grid” and “fabric” layers

Deployment of SAMGrid:
- A gateway node, to install standard middleware from the VDT distribution, the batch adapter, and the client software.
- Worker nodes do not require any pre-installed software or running daemon
Reprocessing DØ data with SAMGrid – AHM 2005

SAMGrid plots

- SAMGrid monitoring
  - Submission and execution sites
  - Jobs monitoring and status
- More than 10 DØ execution sites

http://samgrid.fnal.gov:8080/


-- SAM Grid List of Resources --

<table>
<thead>
<tr>
<th>Station Name</th>
<th>Type</th>
<th>Site</th>
<th>Universe</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>caps10</td>
<td>SAMGRID</td>
<td>LTU</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>cchen2p3-analysis</td>
<td>SAMGRID</td>
<td>cchen2p3</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>cchen2p3-grid1</td>
<td>SAMGRID</td>
<td>LCG_TEST</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>cms-grid</td>
<td>SAMGRID</td>
<td>CMS-FNAL_WC1</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>d0_fcu_prague</td>
<td>SAMGRID</td>
<td>FZU_GRID</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>d0Karlbehe</td>
<td>SAMGRID</td>
<td>GridKa</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>d0ppdg_wisconsin-z</td>
<td>SAMGRID</td>
<td>Wisconsin</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>fnal-farm</td>
<td>SAMGRID</td>
<td>FNAL</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>imperial-prd</td>
<td>SAMGRID</td>
<td>IMPERIAL_PRD</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>lhupack</td>
<td>SAMGRID</td>
<td>LHUPAC</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>oscepa</td>
<td>SAMGRID</td>
<td>OSCER</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>ouhep</td>
<td>SAMGRID</td>
<td>OUHEP</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>samgrid</td>
<td>SAMGRID</td>
<td>Samgrid</td>
<td>prd</td>
<td>cdf</td>
</tr>
<tr>
<td>samgrid</td>
<td>SAMGRID</td>
<td>Samgrid</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>uta-uc</td>
<td>SAMGRID</td>
<td>UTADPCC</td>
<td>prd</td>
<td></td>
</tr>
<tr>
<td>westgrid-ubc</td>
<td>SAMGRID</td>
<td>Westgrid</td>
<td>prd</td>
<td></td>
</tr>
</tbody>
</table>

Please try clearing your browser cache and reloading the page if the information does not seem current.

F. Villeneuve-Seguier @ AHM 2005
DØ data reprocessing

The reconstruction software is significantly improved as the understanding of the detector increases.

- Need to reprocess data frequently with the improved software

Winter 2003/2004
- ~100M events remotely (25M events in UK)
- Distributed computing rather than Grid

April → Oct. 2005 (P17 reprocessing)
- x10 larger, i.e. 1000M events, ~500 TB from raw, DB access
- Basically all remote (more than 10 participating sites)
- SAMGrid based
- 3200 1GHz Pentium III CPU's for 6 months
  - Largest HEP activity on the grid
Site certification and infrastructure

~12 contributing sites to the current data reprocessing
- 7 countries: Brazil, Canada, Czech Rep., France, Germany, UK, USA

Total computing power estimated to > 3500 1GHz Pentium III CPUs.

Each participating site needs to be certified before joining the official production.
- Check for any numerical deviation (configuration/infrastructure problem?)

The certification procedure consists of 2 steps:
- TMB production from RAW data files.
- TMB merging

At each step, many physical variables are plotted and compared with a central reference dataset.
- SAMGrid vs. non SAMGrid
- Site vs. site
Reprocessing DØ data with SAMGrid – AHM 2005

Reprocessing application flow

- RAW data files are stored at the local SAM station cache
- DØ reconstruction software is used to produce TMB files from RAW data files
- TMB files merged (for more efficient use of the database)
- Merged TMB files are stored with SAM to Enstore (FNAL)
Reprocessing DØ data with SAMGrid – AHM 2005

Reprocessing monitoring and status

- Monitoring (illustration)
  - Efficiency, speed by site or overall
  - http://samgrid.fnal.gov:8080/cgi-bin/plot_efficiency.cgi

- SAMGrid enables a common environment & scripts, as well as an effective book-keeping
  - JIM's XML-DB used for monitoring / bug tracing
  - SAM avoids data duplication + defines recovery jobs

- Started end march, and current status is > 85% done, entering the “end-game”.

<table>
<thead>
<tr>
<th>Total Raw Events</th>
<th>986190444</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processed Events</td>
<td>861137054</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sites</th>
<th>finl</th>
<th>FNAL</th>
<th>OSCER</th>
<th>FZU_GRID</th>
<th>WestGrid</th>
<th>ccm2p3</th>
<th>GridKa</th>
<th>UTA-DPCC</th>
<th>Wisconsin</th>
<th>IMPERIAL_PRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS FNAL WC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F. Villeneuve-Seguier @ AHM 2005
SAMGrid interoperability

- Interoperability between different grid projects is a critical issue in the Grid community.
- Ongoing program for SAMGrid interoperability with other grids (in particular with LCG and OSG).
- First step in this direction is the coexistence of SAMGrid on shared resources with other projects.
- Both Monte-Carlo production and the actual data reprocessing are being done at remote sites, using shared resources.
- A significant amount of Monte-Carlo (17M events) has been produced on the LCG by “manual” submission at Nikhef (Netherland).
- Prototype of mechanism to forward jobs from SAMGrid to LCG has been developed and are currently tested in Wuppertal (Germany) and in Lyon (France).
- Prototype being extended, to include 3 UK sites (Imperial College, Lancaster and RAL).
- The SAMGrid/LCG development will “kick-start” the OSG interoperability, as several sites participating in the DØ data reprocessing have OSG resources.
UK Regional Centers

4 UK sites are involved into the SAMGrid project.

More global purpose (Regional Analysis Centers).

Institutes:
- Imperial College and LeSC (London e-science center)
- Lancaster University
- Manchester University
- Rutherford Appleton Laboratory (RAL)

Effort balanced between current reprocessing and Monte-Carlo production.

UK RACs web page: http://www.hep.ph.ic.ac.uk/e-science/projects/d0_rac.html

IC, Lancaster and RAL also joined the SAMGrid/LCG interoperability effort, and started to install and configure the batch adapter prototype tested in France and Germany.
Reprocessing DØ data with SAMGrid – AHM 2005

Conclusion and prospects

The DØ computing model has evolved from a distributed computing model, based on a data grid (SAM), to a computational grid (SAMGrid), based on standard grid tools.

SAMGrid has now been used successfully since winter 2003/2004 for both Monte-Carlo production and for data reprocessing.

Significant advances this year:
- Continued migration to common tools
- Progress on interoperability, both LCG and OSG
  - Successful submission of SAMGrid job on an LCG cluster.
  - Prototype installed and tested at several sites (France, Germany, UK).
- Current data reprocessing
  - Largest HEP grid activity, 1000M events, ~500TB.
  - 12 participating sites from 7 countries, about 3500 1GHz PIII CPUs.
  - Use shared resources.
  - More than 85% complete, expected end mid-october.

Data reprocessing on the Grid is successful, and provides good feedback to the grid community.

A lot more to come soon :)