

CDF Monte Carlo Production Group

Andreas Warburton
McGill

CDF IPP-Canada Workshop
University of Toronto
15 April 2003

Outline

- Some brief recent history of the MCPG
- Broad goals
- The IPP-Canada MC Production proposal
- Current MCPG membership
- Offsite MC production sites
- Web pages
- CDF GRID
- Existing MC requests
- Current problems
- Summer plans status
- To-Do/Priority list
- Mailing list

Recent History of MCPG

(a personal perspective)

- **Heroic push to deliver MC for 2003 Winter Conferences:**
 - described as “successful” by MCPG members themselves
 - Criticisms: coordination lacking, reactive, *in vacua*
 - But: many teething problems have had to be overcome
- **Out of ~31.5M MC events generated for the Winter Conferences, ~8.2M were done at Alberta and Toronto, since Dec. 5th (Bryan and Reda)**
- **runMC executable (see Reda’s talk for more details)**
 - MC-production-specific executable to do simulation and production
 - Something of a kludge
 - Maintained in parallel with Offline releases by Tony Vaiciulis

Broad Goals of CDF MC Production

- **Proactivity**
 - Work within a framework of experiment-wide physics goals
 - Apply these larger goals when prioritizing MC production requests
 - Physics processes
 - Sample sizes
 - Parameter values
- **Reproducibility**
 - Troubleshooting facilitation
- **Reusability**
 - Duplication avoidance
 - Resource conservation (disk, bandwidth, CPU, FTE)
- **Transparency**
 - Documentation
 - TCL script and log-file archives
- **Accountability**
 - Ensure that every MC sample has some DNA mapped to it
- **Strive for Generic CDF MC Sample (difficult!)**
 - à la LEP, CLEO, BaBar/Belle (*i.e.*, e^+e^- machines)

IPP-Canada MC Production Proposal

(Pierre, Reda, AW)

Outline:

- Scope
- MC Production Coordination
- MC Production Group
- IPP Personnel
- IPP Computing Resources
- Status of the IPP Proposal

Scope: Volume and Timescale

(Large Scale Production of Simulation Datasets in Canada)

- We propose to make major contributions to CDF Run-IIa/b data analysis by providing massive off-site computing resources and by developing tools required to facilitate the exploitation of computing facilities at remote institutions.
- Over the next 2 years, we propose that a significant fraction of CDF Monte Carlo simulation datasets be produced in Canada. (>2 years possible)
- Expect capacity of ~1 million ttbar-equivalent events/day (assuming 10 s per ttbar event)
- Note: according to previous planning documents, $\sim 10^9$ events of MC were estimated to be required to serve Run Iia analysis needs
- We would take responsibility for managing our clusters as CDF remote computing resources, monitoring event processing, and transferring the processed data to Fermilab for use by the collaboration.

MC Production Coordination

“We also propose to assume as an IPP institutional responsibility the leadership of the MC production group.”

Charges of the MC production coordinators:

- Work with the Physics Coordinator and the MC Production Group to identify CDF-wide MC needs and priorities in a proactive planning capacity.
- Organize the distribution to off-site MC production sites of tested, standardized, reproducible, and documented tcl files, scripts, and customized executables unique to the MC production process.
- Take responsibility for entering samples into the DFC and onto tape.
- Maintain a detailed web-accessible record of generated datasets and the particulars of their validation.
- Coordinate the implementation of reasonable test procedures to validate the ability of a remote site to produce CDF MC.

MC Production Group Composition/Responsibilities

Composition:

- MC production coordinators
- Physics coordinator
- At least one representative from each physics group

Responsibilities/Charges:

- Work with the Physics Coordinator to identify CDF-wide MC needs and priorities in a proactive planning capacity.
- Coordinate production of samples at participating remote CDF institutes based on physics priorities and resource availability.
- For each new sample, create a unique dataset ID in the DFC (done by the relevant physics group representative).
- Ensure that each generated sample gets validated by its requester.
- Strive for MC sample reproducibility and uniformity across physics groups and samples.

Requirements NOT under the purview of the MC Production Group

- Prioritizing and coordinating MC sample requests within a given physics group.
- Executing and providing documentation for the validation of MC datasets.
- Providing the data handling capabilities at Fermilab for sustained multi-MByte/s data transfers from remote sites.
- Ensuring that the crash rate of the simulation software remains well below 1 in 100,000 events. Assumption:
 - The crash rate of production will continue to be significantly less than that of the simulation.
- Testing and debugging CDF simulation and production codes.
- Providing bug fixes, patches, or new releases of any offline code.

IPP MC Production People

- **Reda and AW:**
 - MC Production coordination
 - Responsible for setting up and running jobs on Toronto and McGill clusters
- **Rachid**
 - Involved in operation and development of Toronto farm?
- **Andrew**
 - Working on issues related to large-scale database usage
- **Jim and Bryan**
 - Coordinate production done on the Alberta cluster
 - Set up and run jobs

IPP Computing Facilities for MC Production

See talks from earlier this morning!

Status of IPP-Canada MC Proposal

- Drafts have been provided to several collaborators:
 - Physics Coordinator
 - A Spoke
 - Current members of the MCPG
 - Some physics-group convenors
- Feedback generally very positive
 - Pierre: “How could anyone refuse?”
 - Several suggestions/improvements applied to draft
 - Proposal scope needs clear definition (“scope bloat”)
- MOU
 - Nigel: get it to Carol
 - But: this proposal doesn’t cover all of the activities planned in the Pierre/Jim/AW grant request (e.g., database, GRID work)
 - Plan: write one MOU describing the complete set of responsibilities

Present MCPG Membership

Physics Coordinator (*ex officio*): Tony Liss

McProd Coordinators: R. Tafirout, A. Warburton

Physics Group Representatives:

- Top: in transition (Tony Vaiciulis departing; replacement identified)
- EWK: **currently nobody** (formerly covered by Top representative)
- Exotics/tau: Martin Hennecke
- Photon: Ray Culbertson
- Bottom: unofficially, Saverio D'Auria (formerly Frank Würthwein)
- QCD: Rick Field

Other Interested Contributors:

- S. Lammel, F. Würthwein, *et al.*
- more welcome...

Off-Site CDF MC Production Sites

(that I'm aware of)

- IPP-Canada
 - Toronto
 - Alberta
 - McGill
- Glasgow: ScotGRID
- Universität Karlsruhe
- UC San Diego
- UIUC
- OSU
- Michigan
- LBNL
- Rutgers
- many more?... (PC husbandry is sexy)

Consolidation of MCPG Web Pages

- Currently have several disparate web spaces devoted to CDF MC production:
 - “Main” MCPG page (authored by Martin Hennecke)
 - Out-of-date
 - IPP-Canada MC Production page (authored and maintained by Reda Tafirout)
 - Up-to-date
 - Exotics 4.9.1 requests page (Martin Hennecke)
 - Up-to-date
 - Top and EWK 4.9.1 dataset requests and inventory page (authored and maintained by Tony Vaiciulis)
 - Up-to-date
- Coordination and centralization of web content is imperative

CDF GRID Project

Principal proponent: Frank Würthwein (FNAL/UCSD)

- Overall scope includes, but extends beyond, MC production
 - User interacts with a standard GUI
 - Job build, deployment, execution, I/O, *etc.* proceeds “under the hood”
 - Geographical invariance: user in Liverpool may have her job execute on the CAF at FNAL, or on Toronto’s Big Mac, or elsewhere
 - but she doesn’t need to care!
- CDF GRID/CAF Software Infrastructure – Main Components:
 - SAM (disk management)
 - JIM (job management, based on Condor and Globus)
 - CAF (interface used by a user for job submission via a GUI)
 - FBSNG (batch management system)
- Issues:
 - How to deal with clusters that aren’t solely devoted to CDF computing
 - Kerberization
 - Compatibility of batch management systems across clusters
- IPP Personnel:
 - Reda, Bryan, Steve, and Pekka are working with Frank on CDF GRID implementation

Present MC Requests

- Temporarily in a lull
 - Winter conferences done
 - Summer offline release coming soon
- Current Outstanding Requests:
 - Exotics group: $Z' \rightarrow ee$
 - “small” sample size
 - TCL under preparation by requesters
 - Exotics (tau) group: 11 decay processes involving tau leptons
 - Sample size $\sim 10.4M$ events
 - Waiting for Big Mac
 - Photon group: large photon samples
 - Pending CES bug fixes (already repaired?)
- Most requesters are waiting for the summer offline release

Problems Currently Known

- **TRIGsim**
 - UIUC farm crashes on Trigsim
 - Alan Sill: Trigsim has known database-access design shortcomings
 - ScotGRID: S. D'Auria reports core dumps using Trigsim to do B MC
- **PlugStripColl issue**
 - Arises when a PlugStripColl object is streamed into and out of an AC++ program without being puffed
 - Large object-size differences observed
 - Eric Brubaker: impact assessment and fix
 - Reda: investigating a temporary fix involving manual puffing via TCL script commands... but requires 4.10 release
- **The UIUC Problem (Trevor Vickey)**
 - Output, as determined from object sizes, looks very different compared with other remote sites
 - Identical TCL and exe configurations compared with other sites
 - Hypothesis: OS configuration (RH and kernel releases)
- **FNAL capability to support sustained data-transfers from remote locations**

Summer Plans

- Physics convenors and MC physics-group representatives have been asked to lay out preliminary summer MC production plans.
- So far: only the photon group (Ray Culbertson) has responded.
- Top representative changing – expect plan soon
- No plan expected from Bottom group this summer
- Exotics, QCD, (EWK?) groups will be prodded this week

List of To-Do's/Priorities

- Firm up group membership (physics-group representatives)
- Normalize web pages
- Establish summer plans and priorities
- Split runMC package into generic executables (Reda, T. Vaiciulis)
- Solve list of currently known problems
- Commission new computing clusters (Toronto, McGill, ...)
- Enable “Direct-to-tape” transfers from remote sites to FNAL (Jeff Tseng, SAM → fstore at FCC)
- Institute validation procedures:
 - Remote site worthiness for MC production
 - Sample-by-sample low-level quality control (diagnostic histo's/ntuples?)
 - Sample-specific physics-quality output validation and documentation (requesters)
- Plan to deal with small-sample requests – Two possibilities:
 - Discourage official requests, but provide How-To's and tools for “self service”
 - Develop and promote simple CAF-like interface
- Standardize B group's approach to MC production

MCPG Listserv

- CDF_MC_Production@fnal.gov
- Archive of CDF MC production traffic is available and searchable
- To date, has not been exploited as much as it should
- Aids in transparency, accountability
- We are maintaining MCPG meeting minutes in this forum

Closing Remarks

- MCPG is undergoing a transition of leadership and organization
- IPP-Canada's MC Production proposal
 - In good shape
 - Well received by Collaboration
- We are behind schedule in summer planning
- IPP-Canada's computing capabilities will play a major role
- Much work lies ahead
- Look forward to achieving a set of very broad goals