May 25 2001
Chlebana

There is a talk at the collaboration meeting in the joint physics session and need to present an overview.

Specify the scope, outline a strategy and write down a functional description of the MC Production project.

Organization

Over the past few weeks we have been talking about developing a plan for generating large MC data sets on the farm.

A very sketchy outline can be found at:

`http://ncdf76.fnal.gov/qcd/mc/`

will be moved to a more official area...

Hopefully this can be filled in with more accurate information and form the basis for a design document.

Email address:

`cdf_mc_production@listserv.fnal.gov`

MC Production Coordinator

Organize the effort...

`Frank Wuerthwein  fkw@fnal.gov`

Physics Group MC Liaisons

Each physics group has identified a MC Liaison.

QCD	Rick Field	rfield@fnal.gov
Top/EWK	Tony Vaiciulis	vaiciuli@fnal.gov
Exotics	Martin Hennecke	hennecke@ekp.physik.uni-karlsruhe.de
B Physics	Frank Wuerthwein	fkw@fnal.gov

Sorting out the requests within the physics group. Look for overlaps between requests. Ensure that the jobs being submitted pass some initial quality checks.

Simulation Coordinator

Contact to ensure that the simulation is up to date.

Define the responsibilities and find a name...

Farm Coordinator

Contact to ensure that events are processed on the farm.

Define the responsibilities and find a name...

Short term goals:

In order to help define the problems we would like to push through a pilot job. This would also be useful in order to evaluate where we are.

Generate 100000 events, write to tape and make available to general users through the data handling system (i.e. AC++ job).

Information can be added to database later on.

Add the functionality to access the data through the data handling system.

Before serious production starts there are issues about default settings that need to be sorted out.

Long term goals:

Split the task into subprojects, provide a more detailed plan (with some implementation details) for each subproject.

Set the priority of each sub project.

Specify the time scale...

Several areas that can be worked on in parallel. We are not necessarily responsible for these all of these points but we will have to be aware of the status and work with contact people...

General Framework

Different people have different preferences on how to generate events. One mode of operation would be to do all steps in one pass.

Event Generation → Detector Simulation → Trigger Simulation → Reconstruction → Ntuple Generation.

In this mode one would write out whatever banks the user is interested in. The ntuple would be immediately available to the user. The simulated data can be reprocessed using different versions of the reconstruction code. New ntuples can be produced.

In the second mode each step could occur in separate jobs. The output of the previous stage could be used as input in the current step.

This is useful for debugging purposes. For example one can use the same generated sample as input to different versions of the simulation.

→ The framework being set up should be flexible enough to handle both modes of operation.

Executable Creation and Parameter Optimization

Could start with cdfSim but would have to add in additional additional modules to provide greater flexibility → add additional generators, trigger simulation...

The configuration settings can be split into separate tcl files which control the distinct parts.

- mc_generator.tcl
  Configures the generator and the settings you want.

- mc_simulation.tcl
  Configures the detector simulation.

- mc_trigSim.tcl
  Performs the trigger simulation.

- mc_reconstruction.tcl
  Production processing

- mc_userModules.tcl
  Configures the user modules (create standard ntuple).

- mc_output.tcl
  Configures the output. Select what banks you want written to the output.

These tcl files can then be swapped in and out as needed. It would be very helpful to have a repository of example tcl files and to identify people responsible for keeping them up to date.

Would like to develop a repository of example tcl files and guide on appropriate settings to use. Henry Frisch expressed some interest in this...

Can learn from the experience of the LLS group.

With the default settings in the simulation it takes 12 sec per event using 3.15.0int2 to simulate HERWIG jet events.

Default parameters in the GEANT simulation are set with low cut off thresholds leading to long execution times. Since the settings may affect the performance of the simulation need to determine a default set that is optimized before tuning the simulation. Input from Physic's groups and simulation contact...

Farm Framework and Data Handling

More details in Frank W's talk...

Statagy for executable to have access to the database.

Handling of random numbers.

May be of interest to provide a facility to look at jobs that crash...

What to do with additional files: log files, ntuples...

Writing datasets to tape.

Entering information in the database: before job starts after job stops.

DataBase and Book Keeping

Keep track of the parameters (tcl files) used to generate the data sets.

Want to use the offline data handling tools and treat MC like data.

Queries into the database to search for datasets and to browse settings used.

User Interface

Interface for users to submit requests.

Survey existing tools used for online.

Quality Control and Monitoring

Avoid having users submitting executables with known bugs.

Small Job Facility

It would be useful to provide a facility to generate small data sets (of a few thousand events).

Should have the ability to store the relevant information in the database and make the data available via the data handling module.

There have been many requests to generate standard MC sets for people to look at.

Generating MC Sets Off Site

More details in Frank W's talk...

Package the code in such a way to make it easy for off site institutes to generate MC sets.

Specify the tape format.

Allow tapes to be shipped to Fermi and input to the data handling system.

Issues of keeping track of random numbers... parameters used for the set.