Computing Summary

XIV SuperB General Meeting
9/27-10/1 LNF

Fabrizio Bianchi
Univ. of Torino & INFN - Torino

Computing sessions in this meeting

- Production
- Open discussion on future activities
- Computing: R&D 1&2

- Det + Comp: Backgrounds
- Det + Phys + Comp: DGWG
- Phys + Comp: FastSim

FastSim improvements

(see D. Brown presentation in FastSim parallel session)

- Neutron interactions modeled in Bruno/G4.
- Pair electrons filter (Svt dE/dx).
- Faster, more accurate EMC time response.
- dE/dx merged hit filtering.
- Detector-based PID selectors.
- Hadronic and Semileptonic Signal cocktails.
- Various low-level code improvements:
  - background frame reading, hit merging, ...
- New navigation.

Old FastSim navigation

- Old model: particle simulation navigation loops over detector elements in a fixed order:
  - order set by configuration.
- Works when particles have a well defined path.
  - outwards through cylindrical shells (SuperB).
- Fails when particles come from unexpected directions, or when elements have no fixed order WRT particle direction.
  - DCH endplate vs backwards EMC.
  - Loss of efficiency near edge of DCH.
  - Missing DIRC info for particles looping to the endplate.

New FastSim Navigation

- Detector space is divided into volumes
- Volumes are divided into voxels
  - cylindrical geometry
  - subdivisions in ρ,Z,φ, defined in configuration

- Voxels reference enclosed detector elements
- Particles are tracked through voxels
  - deterministic
  - independent of any assumptions

- Particles interact with elements within voxels
  - No geometric assumptions about element order inside a voxel.

Production Update

- 147KJob done + 3KJob running : 9.3*10^9 events done so far
- Submission of DG_4 jobs (HAD+SL+Generics, with and without pairs) is finished (100% of initial requests); we expect them to be done by this evening.
- Now submitting DG_4a jobs (HAD+SL, no pairs) to complete request (70% already done)
- DG_BaBar jobs (HAD+SL, no pairs): 20% done.
- Expect to complete requests sometime next week, stay tuned for an announcement.

Next productions

- Current production will be completed sometime next week.
  - After that we'll produce the τ request.
- Please let us know if you need other samples.

Production: lesson learned

- It was a success:
  - Production tools worked properly.
  - Bugs in code found at production time:
    - Wasted CPU time and wasted manpower.
    - Significant FastSim improvements.

- Plans:
  - Consolidate production tools (4 weeks).
  - Enforce validation procedure (1 week).

Production validation: FastSim

- Create an hard link between requests submitter and validation effort.
- Requester should fill the request table including minimum size of validation sample http://mailman.fe.infn.it/superbwiki/index.php/FastSimDoc/Prodseries
- The requester is committed to validate the sample during the pre production period and note the positive or negative result on request table.
  - If validation is not done, the production of the corresponding sample will be stopped.

FullSim plans

- Used by sub-detectors for background and optimization studies.
- Packaging of FullSim.
  - Work started but never completed.
  - Lack of core manpower.
- FullSim production integration in production suite at CNAF to be able to do distributed production(4 week).
- Define and implement a validation stage for FullSim.

Distributed computing

- Production was done in a fully distributed environment.
  - 15 sites have been exploited.

- Can we do the same for analysis ?
  - No analysis on the GRID for the time being
  - Distribute data to some centers and submit analysis jobs were data are.
  - Solution:
    - Data are simultaneously transferred by production jobs to multiple sites
    - Develop a data transfer system
  - Keep full data sample at CNAF

R&D

- Many areas identified at the Ferrara Workshop.
  - Can afford working only on few of them until project is approved.
    - Others are actively investigated by LHC people, can possibly benefit from their work.
- Plan is to:
  - explore, test and deploy building tools.
    - tests of Cmake are already ongoing in Padova.
  - participate in test beds for new distributed storage technologies (Bari, Napoli, ...).
  - refine resource/cost model.
- 1.5 days working meeting in Italy in mid November.

Test of distributed storage (G. Donvito)

Hadoop: few examples

* **Geographical distributed Storage Element**

* Hadoop provides:
  * automatic replica management and storage distribution
  * rack awareness
  * advanced (and *pluggable*) placement policies
  * good monitoring features
* Why don't we try to use it on a WAN environment to see how it works?
  * The concept of rack is used to identify a Site
  * We need a performant WAN link between site
  * It could provide good reliability of data... also in case a whole site become temporarily unavailable

Conclusion

- A lot has been done:
  - Significant improvement to FastSim code.
  - Production requests have been mostly fulfilled.
  - Analysis of produced data is in progress.

- Still a lot to do:
  - New productions.
  - Consolidation of production tools.
  - FullSim integration in production suite.
  - Distributed data access by analysis jobs.
  - R&D program

- More manpower for software developments and computing is needed.