MICE Collaboration

# MCproduction on the grid

Dimitrije Maletic

*Institute of Physics, University of Belgrade*

*MICE collaboration meeting 44, 30th of March 2016.*

# Outline

- MC production on the grid
- MCproduction scripts
- Finished MCproductions on the grid
- Next steps

# About Mcproduction on the grid (1/3)

- The pure Monte Carlo executable script used is **execute_MC.py** in MAUS bin/utilities directory.
- Necessary information for the running of the MC simulation are http/srm link **to list of G4Beamline chunks** (where line number in this list represents the run number), **<SW_Version>**, and **a simulation datacard** details.
  - MAUS accesses the CDB to get appropriate configuration and calibrations, defined by reading of the datacard.
  - Each request/start of MCProduction, will be tagged with unique MCSerialNumber.
  - MCSerialNumber is a row number in CDB, where the simulation datacard, human-readable comment, and the desired <SW_Version> are stored.
- The information needed for the CDB entry is passed to Production manager, who inserts the card into CDB and starts the running of MC production.
- The <some>.sh script, which is passed to the grid by grid job, is used to execute **execute_MC.py** by passing MCSerialNumber and run number.
- MAUS writes a tarball like **<RunNumber>_mc.tar**

# About Mcproduction on the grid (2/3)

- When all jobs are done (~15k), one grid job will get all created **<RunNumber>_mc.tar** and make **<MCSerialNumber>_mc.tar**
  - The **<MCSerialNumber>_mc.tar** is stored on local SE, and possibly to Imperial SE for http access.
  - After the criterion for determining that <MCSerialNumber>_mc.tar has been correctly built is meet, all created (~15k) **<RunNumber>_mc.tar** files will be deleted.

- The file with all produced chunks has a LFN (also has srm and http path)
  **/grid/mice/Simulation/MCproduction/<TenThousands>/<Century>/<MCSerialNumber>_mc.tar**
    - e.g. /grid/mice/Simulation/MCproduction/030000/030100/030101_mc.tar
    - 6 digits implies we will need less than 1e6 MC jobs in the MICE lifetime

- The lfn/guid information of this file is then passed to data mover which replicates the file to castor at RAL PPD.

# About MCproduction on the grid (3/3)

- So, to conclude, the request **for starting of the Mcproduction;** the information which is passed to Production manager, who inserts the card into CDB and starts the running of MC production, are:
  - <SW_Version>
  - http/srm link to list of G4Beamline chunks,
  - datacard details (geometry download details)
  - possibly some comment about MCSimulaton.

- **book-keeping and advertising** (not yet fixed)
  - book-keeping starts with insert of information about resulting tarball into the data mover DB, which also handles the replication of resulting tarball to RAL PPD.
  - Details about MCProduction, inluding MCSerialNumber, http/srm link to list of G4Beamline chunks, <SW_Version> and simulation datacard details can be written on the and advertised to Analysis group/Software group mailing list?

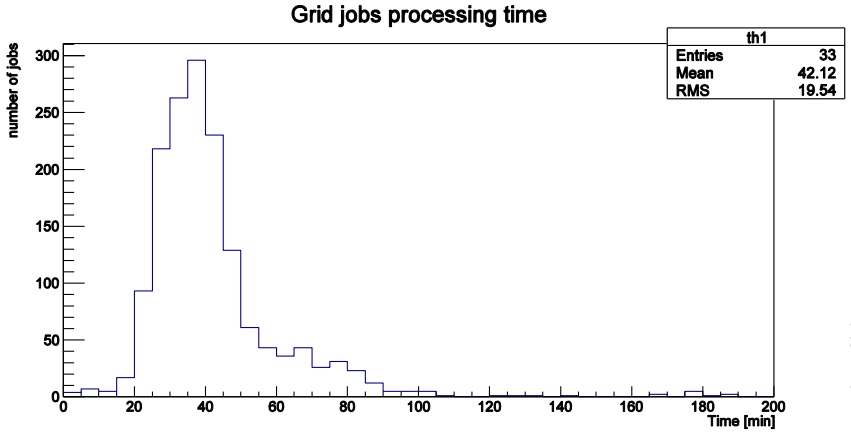
- There is a temporary wiki page on Mice mine about MCproductions:
  - http://micewww.pp.rl.ac.uk/projects/analysis/wiki/MCProduction

# MCproduction scripts

- The scripts used for MC production on the grid are available on launchpad
  - https://launchpad.net/mice-mc-batch-submission
  - The short description would be: MICE MC simulation using g4bl json files as input. Grid jobs submission scripts
    - Scripts helping to submit and monitor status of grid jobs. Developed for MC simulations using G4BeamLine json files as input files for MC simulation in MAUS
  - There is also a detailed README file

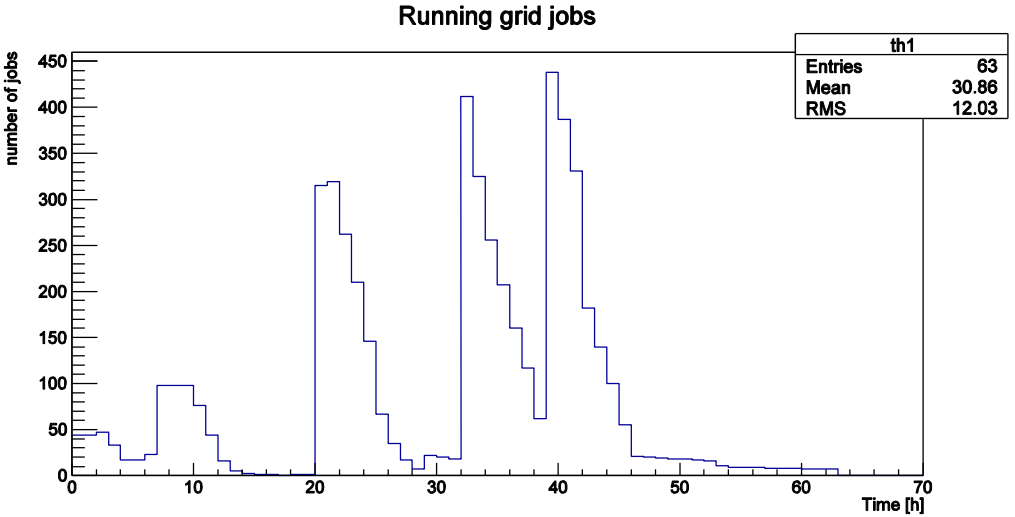
- The submission to the grid and monitoring of the jobs is done using bash scripts and local sqlite db stored at the UI of the submitter.
- The cronjob is set up to start the checking status of grid jobs.
- To create job files and submit them the *create_jdl_and_submit.sh* script is used.
- To check the status of jobs the *check_jobs_db.sh* is used, manually or using the cronjob.
- There are some utilities to *get_run_time_from_std.err_distribution.sh* for done jobs, and a script for merging all chunks to big tarball.

# Finished Mcproductions. Run number 7469.




Grid processing time of each job time in minutes

Number of running grid jobs in time (hours)




## Run number 7469. Status (Done)

MCSerialNumber: 46

comment: Download by run number 7469

softw: 2.0.0

data: geometry_download_by="run_number" geometry_download_run_number=7469

G4BL input

From page Version_3, http://micewww.pp.rl.ac.uk/attachments/download/5568/3200.txt

output: 18GB tarball file with all chunks

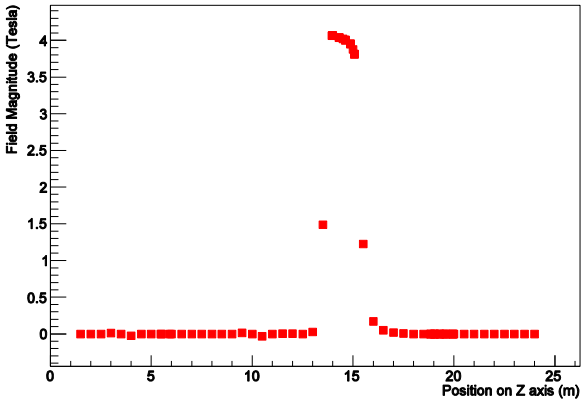
HTTP: http://gfe02.grid.hep.ph.ic.ac.uk:8301/Simulation/MCproduction/000000/000000/000046/000046_mc.tar

LFN: /grid/mice/Simulation/MCproduction/000000/000000/000046/000046_mc.tar

SRM: srm://gfe02.grid.hep.ph.ic.ac.uk/pnfs/hep.ph.ic.ac.uk/data/mice/Simulation/MCproduction/000000/000000/000046/000046_mc.tar
srm://svr018.gla.scotgrid.ac.uk/dpm/gla.scotgrid.ac.uk/home/mice/users/dmaletic/test1/MCProd_out/000046_mc.tar

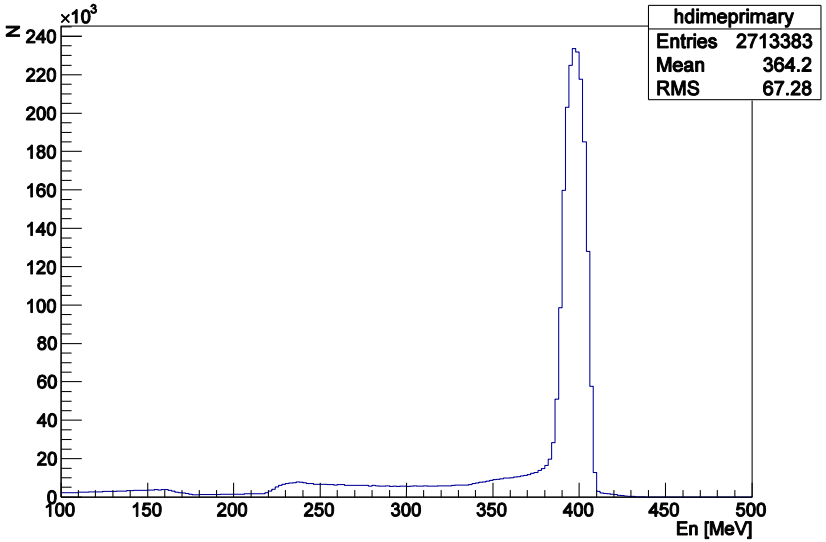
# Finished Mcproductions. Run number 7469.

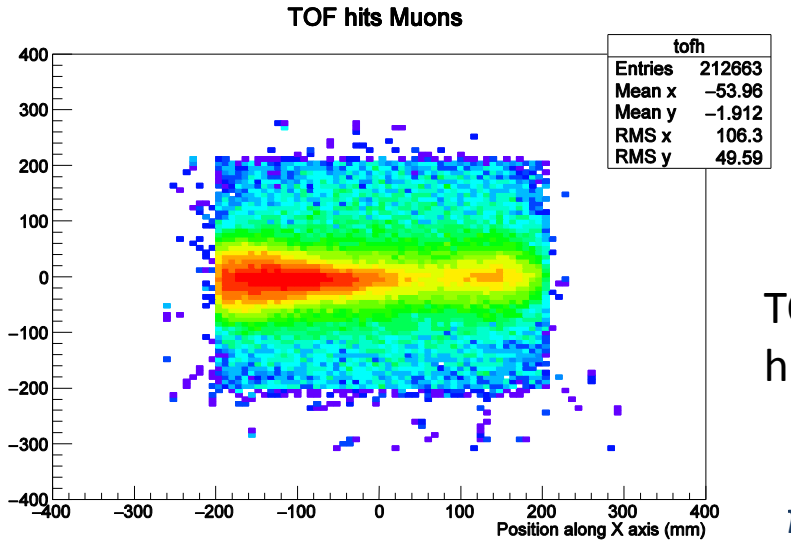
- Preliminary validation of MCproduction
- The list of plots should be discussed further
- Used first 100 chunks
- File name 000046_test_0-99_mc.tar




Magnetic field

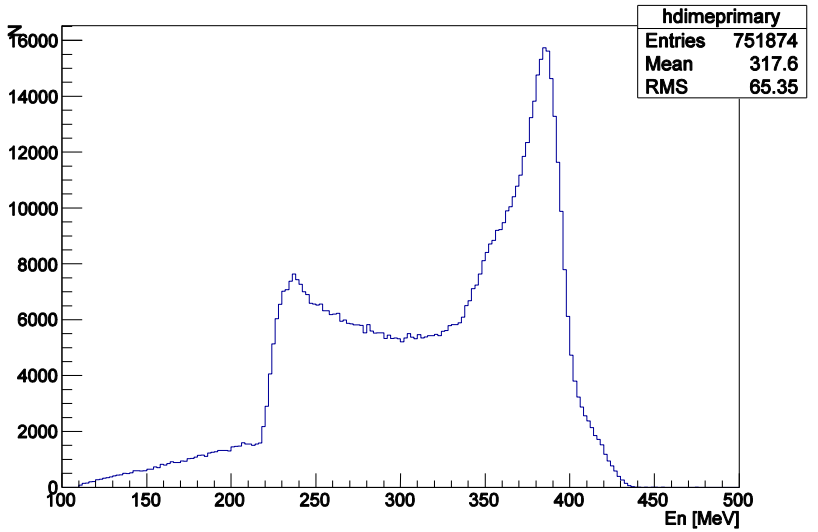
Energy of all primary particles




TOF hits

Energy of muons as primary particles



# Finished Mcproductions. 7417, test geometry 743

**Run number 7417. Note: G4BL 3-200. Status (Done)**

MCSerialNumber: 48. comment: Download by run number 7417 . softw: 2.0.0 . data: geometry_download_by="run_number" geometry_download_run_number=7417

G4BL input list

From page Version_3, http://micewww.pp.rl.ac.uk/attachments/download/5568/3200.txt

output: 18GB tarball file with all chunks

HTTP: http://gfe02.grid.hep.ph.ic.ac.uk:8301/Simulation/MCproduction/000000/000000/000048/000048_mc.tar

LFN: /grid/mice/Simulation/MCproduction/000000/000000/000048/000048_mc.tar

SRM: srm://gfe02.grid.hep.ph.ic.ac.uk/pnfs/hep.ph.ic.ac.uk/data/mice/Simulation/MCproduction/000000/000000/000048/000048_mc.tar
srm://svr018.gla.scotgrid.ac.uk/dpm/gla.scotgrid.ac.uk/home/mice/users/dmaletic/test1/MCProd_out/000048_mc.tar

**Test geometry 743. Note: G4BL 3-200. Status (Done)**

MCSerialNumber: 49

comment: Tag access for 3-200 beam with fields for 200 MeV/c beam. softw: 2.0.0. data:
cdb_download_url="http://preprodcdb.mice.rl.ac.uk/cdb/"
geometry_download_coolingchannel_tag="StepIV-6pi200+solenoid" . geometry_download_beamline_tag="3-200+M0" .
geometry_download_by="id"
geometry_download_id=743

G4BL input list

From page Version_3, http://micewww.pp.rl.ac.uk/attachments/download/5568/3200.txt

output: 2 files: 16GB and 9GB

HTTP: http://gfe02.grid.hep.ph.ic.ac.uk:8301/Simulation/MCproduction/000000/000000/000049/000049_0-999_mc.tar
http://gfe02.grid.hep.ph.ic.ac.uk:8301/Simulation/MCproduction/000000/000000/000049/000049_1000-1563_mc.tar

LFN: /grid/mice/Simulation/MCproduction/000000/000000/000049/000049_0-999_mc.tar
/grid/mice/Simulation/MCproduction/000000/000000/000049/000049_1000-1563_mc.tar

SRM: srm://gfe02.grid.hep.ph.ic.ac.uk/pnfs/hep.ph.ic.ac.uk/data/mice/Simulation/MCproduction/000000/000000/000049/000049_1000-1563_mc.tar
srm://gfe02.grid.hep.ph.ic.ac.uk/pnfs/hep.ph.ic.ac.uk/data/mice/Simulation/MCproduction/000000/000000/000049/000049_0-999_mc.tar
and, all g4bl chunks: srm://svr018.gla.scotgrid.ac.uk/dpm/gla.scotgrid.ac.uk/home/mice/users/dmaletic/test1/MCProd_out/000049_mc.tar

# To do

- Make more precise workflow, especially on
  - MCproduction validation - what to check,
    and if fails what are the next steps (contact persons):
    - Steps towards fixing/correcting of geometry
    - steps towards debugging of the simulation

- Problem of file sizes (if file size is more than limit, grid site administrators thoughts)
- Access for storage on local SEs
- CDB access from MICE hall racks.
- Populate MICE meta DB, data replication to RAL.