

GRID: DC & production tools

P. Bühler

Stefan Meyer Institute for Subatomic Physics, SMI, Vienna

DC & production tools

- Data challenge from 23.-27. November, 09
- Environment for structured data production on the grid

Data challenge from 23.-27. November, 09

- since last DC: update of middleware, new storage scheme
- main purpose was to test performance of Grid
- test case: $\bar{p} + \text{Ca-40}$ at resonance energy of J/ψ with UrQMD
- study invariant mass distribution of reconstructed e^+/e^- pairs
- gather as much statistics as possible

Preparations

- requires UrQMD generator and sim/digi/reco/pid
- wanted to use track fitter with genfit → needed one of the latest pandaroot versions
- urqmd::root520 using panda_extern::jul08 available on grid
- Mohammad prepared package pandaroot::rev6808 using panda_extern::jul09 and installed on all sites
- urqmd::root520 and pandaroot::rev6808 use different external packages → can not be run in the same job

Preparations

- re
- w
- p
- u
- M
- p
- u
- p

```
#
Executable = "/panda/user/p/pbarprod/bin/UrqmdSim_DC03.sh";
Arguments="#alien_counter# $3 $4 $5 $6 $1 $2";
#Requirements=member(other.GridPartitions,"Production");

Packages={
  "pbarprod@panda_extern::jul08",
  "pbarprod@panda_extern::jul09",
  "pbarprod@urqmd::root520",
  "pbarprod@pandaroot::rev6808"
};

InputFile={
  "LF:/panda/user/p/pbarprod/dc03/macros/sim_tpc_DC03.C",
  "LF:/panda/user/p/pbarprod/dc03/macros/digi_tpc_DC03.C",
  "LF:/panda/user/p/pbarprod/dc03/macros/reco_tpc_DC03.C",
  "LF:/panda/user/p/pbarprod/dc03/macros/pid_tpc_DC03.C"
};

OutputFile={
  "Urqmd.root",
  "points_tpccombi.root",
  "params_tpccombi.root",
  "digi_tpccombi.root",
  "analysis.dat"
```

Preparations

- requires UrQMD generator and sim/digi/reco/pid
- wanted to use track fitter with genfit → needed one of the latest pandaroot versions
- urqmd::root520 using panda_extern::jul08 available on grid
- Mohammad prepared package pandaroot::rev6808 using panda_extern::jul09 and installed on all sites
- urqmd::root520 and pandaroot::rev6808 use different external packages → can not be run in the same job
- two types of jobs
 - ▶ jobs **a**: event generator
 - ▶ jobs **b**: doing the rest
- Preparation of a DC03 wiki page
<http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge3>

Execution

- job submission started by hand
- resubmission of failed jobs
- submission of jobs **b** when $\sim 100\%$ of subjobs **a** finished
- update of wiki page by hand

Execution

- jo
- re
- sl
- ul

jdl	runid	nr of split	nr of events/split	momentum [GeV/c]	submit time	MasterJob ID	DONE
Urqmd_DC03	pbar + Ca40						
runUrqmdSim_DC03_a	100	100	100	4.06	23.11.2009, 08:00	779834	100
runUrqmdSim_DC03_b	100	---	100	---	23.11.2009, 08:46	780500	100
runUrqmdSim_DC03_a	101	100	100	4.06	23.11.2009, 08:00	779852	100
runUrqmdSim_DC03_b	101	---	100	---	23.11.2009, 08:52	780601	100
runUrqmdSim_DC03_a	102	100	100	4.06	23.11.2009, 08:00	779892	100
runUrqmdSim_DC03_b	102	---	100	---	23.11.2009, 08:59	780702	100
runUrqmdSim_DC03_a	103	100	100	4.06	23.11.2009, 08:00	779934	99
runUrqmdSim_DC03_b	103	---	100	---	23.11.2009, 10:45	782446	99
runUrqmdSim_DC03_a	104	100	100	4.06	23.11.2009, 08:00	779938	100
runUrqmdSim_DC03_b	104	---	100	---	23.11.2009, 10:21	781843	100
runUrqmdSim_DC03_a	105	100	100	4.06	23.11.2009, 08:38	780398	100
runUrqmdSim_DC03_b	105	---	100	---	23.11.2009, 09:02	780804	100
runUrqmdSim_DC03_a	106	100	100	4.06	23.11.2009, 09:04	780905	100
runUrqmdSim_DC03_b	106	---	100	---	23.11.2009, 10:12	781540	100
runUrqmdSim_DC03_a	107	100	100	4.06	23.11.2009, 09:04	780906	100
runUrqmdSim_DC03_b	107	---	100	---	23.11.2009, 10:19	781742	98
runUrqmdSim_DC03_a	108	100	100	4.06	23.11.2009, 09:04	780907	100
runUrqmdSim_DC03_b	108	---	100	---	23.11.2009, 10:03	781338	100
runUrqmdSim_DC03_a	109	100	100	4.06	23.11.2009, 09:04	780908	100
runUrqmdSim_DC03_b	109	---	100	---	23.11.2009, 10:17	781641	100

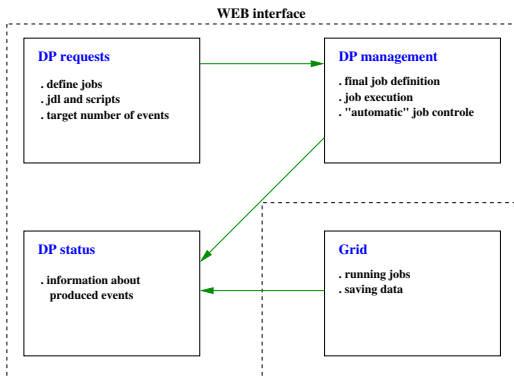
Execution

- increased number of splits per job to 500/1000
- 100 events per split → 170 MB of output
- many jobs simultaneously in status SAVING → system started to become unstable
- e.g. new jobs were only accepted after restart of central services

Lessons learned

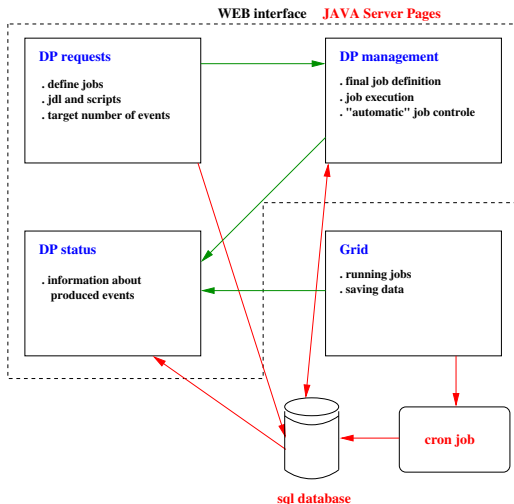
- 9.0 M **a**-events, 2.27 M **b**-events DONE
- SAVING data from worker nodes to SE is critical
- appropriate tools to measure SE performance need to be put in place
- → allows to detect/remove bottlenecks
- reco frequently aborted → stderr can be provided for further analysis

Environment for structured data production on the grid



LPM - Lazy Data Manager

- system developed for ALICE



PANDA's LPM

- part of PANDA Grid monitoring site at <http://mlr2.gla.ac.uk:7001/>
- needs further adaption
- standardisation for request definition required → which parameters, files are needed to completely define a task, e.g.
 - ▶ *jdl(pars)*, *simexe.sh*, *gen.C*, *sim.C*, *digi.C*, ...
 - ▶ *jdl*, *pandaexe.sh*, *pandaexe.conf*
 - ▶ *pars*
 - ▶ ...

Next steps

- analyze bottlenecks and investigate improvements
- adjust/optimize LPM for PANDA
- the PANDA grid is up-and-running and should be used for concerted simulation campaigns