Subject: Upcoming DC
Posted by Johan Messchendorp on Wed, 02 Apr 2008 20:01:24 GMT
View Forum Message <> Reply to Message

Hi all,

Let me start a forum discussion concerning the data production test on the Grid. From the email correspondences, I would conclude the following (please correct me if I am wrong):

  * Suggestion by Kilian to have REGULAR production tests on the GRID is welcomed (test performances, stability, capacity, monitoring tools, etc etc).

  * Proposal is to have it at least twice per year, preferably connected to a Grid workshop.

  * Kick-off of such DS now! Don't wait for something better to come.

Now the critical issue: which jobs? My proposal:

  1) the default QA macros as used in the nightly tests (Dashboard). At the moment there is not a lot of them, but certainly this will be extended during course of time. Actually, these macros are meant as rigorous tests (simulation, reconstruction, analysis) of the framework, e.g. why not exploit these macros on the Grid sites as well. It would also garuantee that the software developments are tested in synchronization with the Grid infrastructure. There will also be a "master" macro which calls the individual QA macros. All these macros will be part of the PandaRoot software.

  2) software compilation and building. As we already discussed via email, the preference for the moment is to employ the PandaRoot software via a friendly installation script. Mohammad and Florian are already doing excellent work on this front and working hard on making the script as grid-friendly as possible. A regular compilation and building on all the sites would reveal missing configuration issues, but would also give regular feedback to Mohammad and Florian. One should realize that sites might also change their installation during course of time, which would lead to crashing jobs for which the problem is difficult to trace back. A regular compilation test would help to identify and accomodate for these kind of changes.

.... what else?

Johan.

Subject: Re: Upcoming DC
Posted by Kilian Schwarz on Thu, 03 Apr 2008 06:50:20 GMT
View Forum Message <> Reply to Message

I support the idea to run the QA macros on the Grid. Also to start the DC#1 now. Do we take the dates suggested by Dan ?
He suggests to start on April 17 and run for 3 days.
Also the one with the regular builds on Grid sites with feedback to the developers is a good idea, I think.

Cheers,

Kilian

Subject: Re: Data Challenge
Posted by Dan Protopopescu on Thu, 03 Apr 2008 10:34:42 GMT

View Forum Message <> Reply to Message

Dear all,

I would like to thank everybody for participating in this discussion. There are several that showed up and I will address herein:

1) Dates of the DCs
2) Nightly builds on Grid
3) Proposed test jobs/macros
4) Storage of the outputs

Here below is my opinion on these issues. Please use the forum to reply so that we have an organized thread. If you have no access, register or ask someone else to post for you.

1) The Grid Data Challenges will happen at dates to be decided independently and to ensure this objectivity I propose that the dates be set by our Production Manager (PM), Paul Buehler, after some consultation with both the Grid Coordinator (Dan) and Software Development Coordinator (Johan). The next date is April 17 (no changes accepted) but after that I hope the proposed PM scheme will be applied.

2) One of the major points to be understood about Grid is that it is not a testbench. It is a "massive computations infrastructure", designed for large-scale, stadardized jobs. Although nightly builds on several platforms is an extremely useful tool for developers to spot early compilation problems, it would be a misuse of the grid. The software installed on the grid is supposed to be a stable version, which has been tested already on the platforms existent on the grid sites. Testing and feedback would happen at installation time. Florian explained this issue better than me.

3) The basic plan for the next data challenge is to count jobs and produce statistics like job success rate, job site distribution, time per 1000 jobs. I propose the tests to done both with a generic job and a PandaRoot job in order to decouple the various requirements. Having a macro that simulates some tracks in the EMC or producing some real physics is almost irrelevant for this data challenge. However, I enthusistically embrace the idea of running some physics from which someone can collect, verify the results and gain an extra benefit. If you have such macro, let's use it!
My initial plan is this:
- 10x100 subjobs generic (site availability)
- 1x1000 subjobs generic (broker optimization)
- 1x1000 PandaROOT macro #1
- 1x1000 PandaROOT macro #2
- 1x1000 PandaROOT macro #1 or #2 with all output to Glasgow SE
- 1x1000 PandaROOT macro #1 or #2 with output to local SEs
Please feel free to add to this list and let's discuss the benefits of these tests.
Package testing can not be part of the data challenge because of time constraints. A check of

---

the software installation is part of the preparations taken individually by the site admins.

4) The outputs of the DC test jobs will go physically to local and central SEs and that is part of the challenge. In the file catalogue, the output can be collected as you see fit, in case you would like to use the results later.

Cheers,
Dan

Subject: Re: Data Challenge
Posted by Kilian Schwarz on Thu, 03 Apr 2008 10:46:55 GMT
View Forum Message <> Reply to Message

Hi Dan,

concerning point 2):
the idea was not to use Grid as a test bench. We know very well that software which has been carefully tested locally on various platforms still fails in compiling on various sites simply because something changed there or something is missing or wrong. Therefore such a regular build would not be a test of the software of which we know that it should compile but rather a test of the sanity of the various sites. And from this point of view for my feeling it would make sense.
Also it would give us the certainty that at least the software packages are available and running on the sites. We will have enough other problems during the upcoming DCs.

1) ok

3) ok, but if possible let us try to include some reasonable and meaningful PandaRoot macros if we have. Either the QA or the suggested fast sim macros or both.
SEs: how much output do we expect alltogether and how much storage capacity do we have in the Grid and at the sites ?

Cheers,

Kilian

Subject: Re: Data Challenge
Posted by Dan Protopopescu on Thu, 03 Apr 2008 11:49:44 GMT
View Forum Message <> Reply to Message

I suggest to move the 'Regular Builds' issue on a different thread, because it should not be connected to the upcoming DC:
  http://forum.gsi.de/index.php?t=tree&th=1824&start=0&rid=104
6&S=7da716fd1afe83d0620759523c2e93a7

About your question on how much output do we expect from the test jobs, do you expect me to know that before we even decided what macro to run ? My name is Protopopescu, not Nostradamus.

Subject: Re: Data Challenge
Posted by Kilian Schwarz on Thu, 03 Apr 2008 13:43:01 GMT
View Forum Message <> Reply to Message

Hi Nostradamus,

I did not expect you to know the answer. I just wanted to make aware that we have to be aware of the storage capacity we have and we have to have a plan at least concerning how much we need.

Cheers,

Kilian

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Fri, 04 Apr 2008 11:49:16 GMT
View Forum Message <> Reply to Message

Is there a new version of the software that you would like distributed and installed prior to the next DC ? Software compilation testing can
not be part of the DC itself because it is a asycronous process and would take too long (at least for the upcoming DC).

Let us also talk about the macros you would like to propose for the upcoming DC and what data would you like to keep and where.

Subject: Re: Upcoming DC
Posted by Paul Buehler on Fri, 04 Apr 2008 12:51:10 GMT
View Forum Message <> Reply to Message

Dear all,

For this first DC I think the programs do not need to produce meaningful data. This will certainly come in the following DCs. What could be interesting is to have a script which allows to specify how long a program will be running on a site and how many bytes of output it will produce.

With such a program we could make different specific tests, like

submit a large amount of short lasting jobs to test how many of them are successfully finished or fewer jobs producing huge data files to test more specifically the data transfer capabilities, or ...

Preferably this program is a PANDARoot application but not necessarily.

Any comments on that?

Regards,

Paul

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Fri, 04 Apr 2008 13:02:07 GMT
View Forum Message <> Reply to Message

Hi,

This weekend, I will look into a few macros which would be useful to test the computing capacity and data transfer. I will try to be creative. Florian, would you have time to install the new fairsoft stuff on the Grid via Packman?

I would really insist on testing the system using pandaroot applications. I think this is also important from a political point of view, since it would show that the software and the computing infrastructure are developed hand-in-hand. We should certainly avoid a situation in which the software is not compatible with the infrastructure (or the other way around). I agree, in principle this should not happen, but in reality it can, as we have seen with the old framework.

Kind wishes,

Johan.

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Fri, 04 Apr 2008 13:49:27 GMT
View Forum Message <> Reply to Message

We could have both: a generic job as well as PandaRoot macros producing meaningful data. This way we can practically decouple these components for the evaluation stage. I would encourage further discussions about both options.

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Fri, 04 Apr 2008 14:31:02 GMT
View Forum Message <> Reply to Message

Hi,

Can we have a short EVO meeting to discuss the preparations for the DC?

Johan.

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Fri, 04 Apr 2008 14:40:01 GMT

Wouldn't the forum be preferable because then we have a written plan of action ? There is not too much actually, is it ?

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Fri, 04 Apr 2008 14:51:27 GMT

Hi Dan,

I certainly wouldn't mind to stick to the forum. But is it clear, who is preparing what?

Johan.

Subject: Re: Data Challenge
Posted by Johan Messchendorp on Fri, 04 Apr 2008 14:55:23 GMT

Maybe we should turn the question around: how much data do we want to produce for the DC? We can easily tune the macros for that (i.e. decide on how many events to throw for example)

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Fri, 04 Apr 2008 15:21:42 GMT

My initial plan as outlined earlier would be this:
- 10x100 subjobs generic (site availability)
- 1x1000 subjobs generic (job broker evaluation)
- 1x1000 PandaROOT macro #1
- 1x1000 PandaROOT macro #2
- 1x1000 PandaROOT macro #1 or #2 with all output to Glasgow SE
- 1x1000 PandaROOT macro #1 or #2 with output to local SEs
Package testing not be part of the data challenge because of time constraints.

I think the first two list items coincide with what Paul proposed but we should refine our choice of 'generic' jobs.

Could Johan and Soeren decide on the two macros ? Let's say one on detector simualtions and one conatining physics. We already tested the fast simulations macro (Johan), and Soeren proposed including some rho analysis.

Please feel free to add to the above list and we'll decide next week on the final set.

What do you think about such a start ?

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Fri, 04 Apr 2008 15:43:29 GMT
View Forum Message <> Reply to Message

No problem. Soeren, do you have any preference concerning the physics channel we should run? Here would be my suggestion: we could run for instance eta_c channel to multi-photon channels using fast simulations and rho, and run at the same time a full MC simulation with the same eventgenerator output to make a benchmark/validation for the fast simulations (G3 versus G4 versus fast simulations). Alternatively (or in addition), we could consider to run a channel with many charged pions in combination with conformal mapping code (MVD+TPC).

Johan.

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Fri, 04 Apr 2008 17:36:41 GMT
View Forum Message <> Reply to Message

I created a wiki to list the things to do, and where we can attach macros, JDLs and later tables and figures with the results:
http://panda-wiki.gsi.de/cgi-bin/viewauth/Computing/DataChallenge1
What do you think ?

Subject: Re: Upcoming DC
Posted by Kilian Schwarz on Sat, 05 Apr 2008 06:11:32 GMT
View Forum Message <> Reply to Message

Hi Dan,

nicely done. Now we only have to fill it with live.
Should we put there the fast sim macros ?
This would have the advantage that we would also exercise the CERNlib package installed by Johan. At least if we produced everything on the sites.

Cheers,

Kilian

Subject: Re: Data Challenge
Posted by Kilian Schwarz on Sat, 05 Apr 2008 07:08:39 GMT
View Forum Message <> Reply to Message

yes, this is a good one. Not only the number of jobs but also the amount of produced data is an important parameter

Cheers,

Kilian

Subject: Re: Upcoming DC
Posted by Kilian Schwarz on Sat, 05 Apr 2008 07:10:09 GMT
View Forum Message <> Reply to Message

yes, of course. When ?

Cheers,

Kilian

Subject: Re: Upcoming DC
Posted by Jens Sören Lange   on Sun, 06 Apr 2008 13:28:19 GMT
View Forum Message <> Reply to Message

Hi all,

Stefano and I thought a bit again about the question,
which macros to use.

And the key question is:

****
which are our most time-consuming steps in the simu or reco?
****

And, actually, here the fast sim is the smallest problem,
because - obviously - it is fast by definition. So I changed my opinion a bit.

Therefore I would like to propose three different macro groups
(in other words, our "bottlenecks")

1.) dpm
2.) UrQMD
3.) tpc reco and stt reco
(maybe - if we keep the DC data somewhere - we could actually use it for the long-planned
tpc/stt comparison?)

So, concerning 1.)

macro/run/run_sim1.C
with all detectors switched on
and then change the generator to DPM, see
http://panda-wiki.gsi.de/cgi-bin/view/Computing/Dpm
-> "Simulation inside PandaRoot"
and then generate billions of events

Note: unfortunately I don't know anymore how to set the beam energy in
DPM. I have to ask Stefano tomorrow.

So, concerning 2.)

macro/run/run_sim1.C
with all detectors switched on
and then change the generator to UrQMD, see
http://panda-wiki.gsi.de/cgi-bin/view/Computing/UrqmdSmm
here the heavy targets (Au, Pb) are most useful,
because most time-consuming.
anti-proton beam momenta 3.00 and 4.05 GeV
(these are needed for the J/Psi-in-nucleus measurement).
Actually, the GRID would be very useful here to generate as many events as possible!

So, concerning 3.)

macro/tpc/tutorial
runMC.C runDigi.C runReco.C
the svn version of these macros have some difficulties right now
(I just tried again and I have e.g. undefined symbol GeaneTrackRep),
see also
 http://forum.gsi.de/index.php?t=msg&th=1802&rid=0&S=dfa54395
2d09c2dca876d4fb1bde7c98#msg_6124
e.g. one has to comment out "UseGeane()".
I hope that we can fix it until the DC
(I know that Dipak has a version which works).

macro/stt
run.C rundigi.C runreco.C
they work fine.

Here I would propose just to use the box generator
for muons with
pT=30,40,50,...100 MeV/c
pT=100,200,300,...,1000 MeV/c
pT=1,2,3,...,7.5 GeV/c
and uniform polar angle.
(the highest point at 7.5 GeV/c is for the Drell-Yan measurement).

What do you think?

cheers, Soeren

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Mon, 07 Apr 2008 10:20:53 GMT
View Forum Message <> Reply to Message

I am very glad to see that we have interesting stuff to run. And yes, we can run all these
macros. I propose that we prepare as follows:
1) make sure the macros run (on your desktop) -> Soeren, Stefano ?
2) make sure we have the latest software tarballs (including all the updated software) ->
Florian, Johan, Mohammad ? + all SA

3) make an estimate of the number of jobs/events you wish to run -> Soeren ?

With regard to (2), we will need the support of all site admins (SA), to install and check the installation of the new package version once it is made available. We have to keep in touch.

The location where the output goes is set in the JDL: what to keep, where to register it, where to save it physically.

The macros themselves can be updated in the last moment, then added to the alien catalogue etc.

The related wiki topic is: http://nuclear.gla.ac.uk/twiki/bin/view.pl/Main/SubmitExample.

Please document your contributions to the upcoming DC in:
http://panda-wiki.gsi.de/cgi-bin/viewauth/Computing/DataChallenge1

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Mon, 07 Apr 2008 12:33:21 GMT
View Forum Message <> Reply to Message

Hi Soeren,

Actually, how many UrQMD events do we have on disk? I had the impression from Alicia that the available statistics is very poor, right? The question is whether one could run the event generator on the Grid machines to produce more data.

Kind wishes,

Johan.

Subject: Re: Upcoming DC
Posted by asanchez on Mon, 07 Apr 2008 13:00:33 GMT
View Forum Message <> Reply to Message

Hi Johan
Aida has created more urqmd+smm files
simply take a look into /panda02/urqmd_smm
but these results are not related with pure urqmd
she has added one file more for each energy.

best regrads
Alicia

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Mon, 07 Apr 2008 13:38:20 GMT

Hi Alicia,

Ok! Do you know roughly how many events in total there are?

Johan.

Subject: Re: Upcoming DC
Posted by asanchez on Mon, 07 Apr 2008 13:44:28 GMT

HI
So far i remember
she did one million.

Subject: Re: Upcoming DC
Posted by asanchez on Mon, 07 Apr 2008 13:49:02 GMT

Hi here i posted

the Aida's email

so good luck

cheers
ALicia.

 Hi all, hi Alicia,

I have calculated 1 million Events
of pbar-C12 interactions at 3 GeV/c momentum using Urqmd_SMM
generator.

The calculations are disposed at lxi00X.gsi.de cluster
in directory /d/panda02/urqmd_smm/pbarC_3GeV
in 10 files: pbarC_1.root, pbarC_2.root ... pbarC_10.root .
Each file contains 100000 Events of pbar-C Events.

Good luck,
Aida

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Mon, 07 Apr 2008 13:59:08 GMT

Thanks Alicia,

Actually, do you need more statistics for your analysis? Such a DC, might be useful for that....

Johan.

Subject: Re: Upcoming DC
Posted by Jens Sören Lange on Mon, 07 Apr 2008 14:58:45 GMT

Hi Dan and all

... what is the deadline for the macros?

(unfortunately semester started last week which means the days are already almost filled even without Panda ...)

Soeren

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Mon, 07 Apr 2008 15:08:35 GMT

During the DC, we will basically run what will be provided to us by April 15, plus some generic benchmark jobs.

Of course, if you have some physics to run, it can be run outside the DC. The Grid is available in general and everyone is welcome!

Subject: Re: Upcoming DC
Posted by Jens Sören Lange on Mon, 07 Apr 2008 16:35:33 GMT

Hi Dan,

OK it means we got a few more days.

The point is:

Stefano just showed me that there is a chance (he even has a running macro already) to have running tpc sim and tpc digi+reco up to clustering (and probably mvd digi+reco, but here he still needs some help from Ralf or Tobias) entering into lhetrack and maybe even riemannfit (that means we would have digis and even tracks in the data!) and that for many different

values for pT. I don't have to say that this would be a jackpot. However, it needs some more days. And: no promises. It could be that in the end it is not running and we have to stop at the tpc clusters (but anyway, that data set would be great - usually I ran tpc clustering for 1,000 tracks in 3-4 hours).

The bad news:

we found that it is (so far) not possible to run dpm or UrQMD from a macro (only read the data files from a macro).

Anyway, we need help in figuring out how we change parameters in the macros (e.g. changing the pT) from outside when you submit the job ...

cheers, Soeren

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Mon, 07 Apr 2008 16:49:48 GMT

Hi all,

To run DPM is not so difficult, since we anyway run from a (bash) script. The jdl scripts can take arguments, which can be passed to the shell scripts. Actually, last time we run the DPM generator in combination with fast simulations.

Just make sure that also the macros use input parameters (random number see, input filename, outputfilename, energy,..), then it is very trivial (see example below).

(ps, we might have to think a little bit more about the random number seed. I am not sure whether one can take any number for that)

Johan

-- example shell scripts called by JDL --

```
#!/bin/bash
#
# $1 unique number for seed
# $2 momentum
# $3 number of events
#
#
```

```
echo "This is the fast simulation test production"
echo "provided by Johan Messchendorp"

echo "Starting the job"

export RANSEED=`expr 1202677345 + $1`

cat <<EOF >input.$1
$RANSEED
$2
1
$3
EOF

cat input.$1
DPMGen < input.$1

root -b -q "simfast_jgm.C(\"Background-micro.root\",0,$3,\"simfast_jgm.root\")" ||
exit 11
echo "------------------------------------------------------------------ "
echo "From wrapper script: job finished successfully"
```

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Mon, 07 Apr 2008 20:53:48 GMT
View Forum Message <> Reply to Message

Dear all,

I installed on the Grid (which means via PackMan)

pandaroot rev2432 (7/4/08)
dpmgen rev2432 (which means derived and build from the pandaroot rev2432)

against

cbmsoft 16/01/08 (geant4.9.1, geant4_vmc_r331, geant3.1.9, vgm 3.00, root5.18, pluto412, pythia6, clhep2.0.3_1).

If we want to run Geant4 simulations in the DC, it is probably advisable to upgrade as well the cbmsoft release to the most recent one of march'08. Otherwise, I would say, lets stick to this for the upcoming DC event, since cbmsoft_r16/01/08 already compiled successfully on many/most of the sites. Please note, that I only compiled the above pandaroot-related packages on the KVI site. I expect, however, no problems in the compilation of PandaRoot on the other sites. But that I leave up to the site administrators to test:

packman install pandaroot::rev2432
packman install dpmgen::rev2432

(probably, "packman install dpmgen::rev2432" will do the job since it depends on pandaroot)

After installation, one can test it by (existing scripts from last runs)
(dpmgen, 5.5 GeV/c, 10x1000 events, in combi with fsim)

submit /panda/user/p/pbarprod/jdl/simfast_jgm.jdl 999 5.5 1000

with the output written to

/panda/user/p/pbarprod/jgm/fast/run999/1-10

Johan.

Subject: Re: Upcoming DC
Posted by Kilian Schwarz on Tue, 08 Apr 2008 14:42:05 GMT
View Forum Message <> Reply to Message

I installed all packages at GSI and they seem to work.
Testjob worked, too.

Cheers,

Kilian

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Tue, 08 Apr 2008 15:56:05 GMT
View Forum Message <> Reply to Message

Thank you very much, Johan, for installing and testing the packages. I will test them on my sites too.
I started a filling the table in:
http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1#Packages
Cheers. Dan

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Tue, 08 Apr 2008 15:57:38 GMT
View Forum Message <> Reply to Message

Thank you very much, Kilian, for testing the packages at GSI.
I started a filling the table in (you might want to take a look):
http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1#Packages
Cheers. Dan

Subject: Re: Upcoming DC
Posted by Johan Messchendorp on Tue, 08 Apr 2008 22:06:51 GMT

Dear Dan and others,

This afternoon (oeps, looking at the clock meanwhile yesterday), we (Soeren, Mohammad, Kilian, and me) discussed briefly after computing model meeting about what to get ready for the DC. Below you find some of the points discussed. The "software" deadline is at the end of the week (i.e. software should be ready on the Grid during the weekend)

1) we will install the new external packages with the "grid" compilation option (should also be more G4 friendly and installation friendlier)

2) we will also put the latest pandaroot release, since there are some new updates from the detector people which might be useful to have for the DC

3) hopefully the macros will be ready at the end of week. I will make the corresponding scripts for the Grid production. We will make as well an estimate on the datasize needed.

4) not yet clear whether we will have the UrQMD eventgenerator available for the DC. We have to see on this. Anyway, email exchanges with Adai and Vladimir are ongoing.

Did I forget anything important?

Johan.

Subject: Re: Upcoming DC
Posted by Kilian Schwarz on Wed, 09 Apr 2008 04:52:38 GMT

Hi Johan,

yes:
- Soeren will check the output generated by the various macros and make a stastistic estimate and based on that a guess how many data we would produce altogether during the DC
- Dan can compare that with the various storage capacities we have and ring the alarm bell if this should not be sufficient
- I will try to find more SE capacities at GSI for Panda. CPU capacities will be more than enough, I guess.

Cheers,

Kilian

Subject: Re: Upcoming DC
Posted by Jens Sören Lange on Wed, 09 Apr 2008 14:07:12 GMT

Hi all,

the stt and tpc macros for the DC are now checked in.

pandaroot/macro/dc1

NOTE that infile and outfile are still fixed in the macros.

stt
===

root -b runsim.C"(nEvents,pT)"
root -b rundigi.C
root -b runreco.C

params 1175k+11.7k/event (params really seem to be run dependand)
sim 240k+6.5k/event
digi 10k+10.6k/event
reco 0.3k+0.2k/event

differences to usual svn macros:

o field maps instead of constant field
o nEvents and pT are option parsed in sim
o nEvents is zero (=read all) in digi and reco

Proposal:
pT=30,40,50,...100 MeV/c
pT=100,200,300,...,1000 MeV/c
pT=1,2,3,...,7.5 GeV/c
10,000,000 events each

tpc
===

we will use Stefano's new macros

root -b run_sim_tpcmvd.C"(nEvents,pT)"
root -b run_rectrack_tpcmvd.C

params   43k (params seem to be fixed)
sim   46k+17.4k/event
digi+reco   zero kB+39.2k/event

differences to usual svn macros:

o nEvents and pT are option parsed in sim

Proposal:
pT=30,40,50,...100 MeV/c
pT=100,200,300,...,1000 MeV/c
pT=1,2,3,...,7.5 GeV/c

10,000,000 events each

dpm
===

not macros, but here Johan's bash scripts will be used.

however, for Kilian's calculation of disk space:

250 bytes/event

nEvents and beam momenta for dpm will be proposed tomorrow.

cheers, Soeren

Subject: Re: Upcoming DC
Posted by Dan Protopopescu on Wed, 09 Apr 2008 14:21:06 GMT
View Forum Message <> Reply to Message

Dear Johan, Kilian and all,

Thank you for all the work put into this. The outputs that we want to keep for later analysis could be directed to one of big SE (I have at least 1T in Glasgow).

We should not worry about the CPU resources; one of the goals of this DC is to see how much is actually available. It will depend a lot on the site admins and the way they allocate resources.

I understand that everyone should wait until Monday to install the packages on their sites. I hope the site admins follow this forum.

I will be travelling on April 14 and 15, and be back on the 16th.

Subject: Re: Upcoming DC
Posted by Jens Sören Lange on Wed, 09 Apr 2008 14:45:59 GMT
View Forum Message <> Reply to Message

Hi all,

please wait a bit before checking out the svn.

I have overlooked that I have hidden svn files in the directories (because I copied whole directories before) so when committing the svn overwrote _other_ directories (!). Mohammad is trying to fix it right now.

apologies, Soeren

Subject: Test Jobs
Posted by Dan Protopopescu on Wed, 09 Apr 2008 15:06:55 GMT
View Forum Message <> Reply to Message

Dear all,

In the test job JDLs please make sure they save output (if the case) to a directory writable by the production user 'pbarprod',
even if you try the job as yourself or as 'pbartest'. We will use 'pbarprod' for the data challenge next week.

Of course, I will doublecheck the JDLs proposed for the DC.

Subject: Informative Titles
Posted by Dan Protopopescu on Wed, 09 Apr 2008 15:09:18 GMT
View Forum Message <> Reply to Message

I also propose to use for our posts more informative titles than the default 'Re: Upcoming DC'.

Subject: tpc and stt DC macros are in svn
Posted by Jens Sören Lange   on Wed, 09 Apr 2008 15:16:27 GMT
View Forum Message <> Reply to Message

rev2450 contains now pandaroot/macro/dc1

Subject: UrQMD_Smm
Posted by Johan Messchendorp on Wed, 09 Apr 2008 20:01:54 GMT
View Forum Message <> Reply to Message

Hi,

I installed a 32 bit and a 64 bit version (binaries with libs only) of the UrQMDSmm
eventgenerator on the Grid, compiled against root5.18 (which version is the most actual one
for the external packages). It seems to work for the sites I could test it on. To test it on any
site:

Installation (presently already instaled at KVI,GSI,Juelich,Dubna,Bucharest)
----
packman install pbarprod@urqmd::r2008

test scripts

/panda/user/p/pbarprod/bin/testurqmd.sh
/panda/user/p/pbarprod/jdl/testurqmd.jdl

i.e.

submit /panda/user/p/pbarprod/jdl/testurqmd.jdl <id> <momentum> <nrofevents>

will do the job with the output @

/panda/user/p/pbarprod/jgm/urqmd/run<id>/..

Johan.
(ps thanks to Aida and Vladimir!)

Subject: Storage@KVI
Posted by Johan Messchendorp on Wed, 09 Apr 2008 20:13:18 GMT
View Forum Message <> Reply to Message

Hi,

I have made 2TByte storage space (temporarily) available at KVI in case needed for the DC.
This space should already be accessible via the SE.

Johan.

Subject: Pledged resources on our DC01 wiki
Posted by Dan Protopopescu on Thu, 10 Apr 2008 13:02:33 GMT
View Forum Message <> Reply to Message

Dear all,
I added tables with pledged resources on our wiki:
http://panda-wiki.gsi.de/cgi-bin/viewauth/Computing/DataChallenge1
Could you please help update these numbers for all the sites ?
Could we also log data on the package installation results ?

Subject: one more week left
Posted by Paul Buehler on Thu, 10 Apr 2008 14:11:01 GMT
View Forum Message <> Reply to Message

Dear colleagues,

I prepared a schedule for the days remaining until the start of the Data Challenge
(summarizing what has been discussed so far). It is displayed on
http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1#Schedule. Inputs, comments
are highly appreciated. I will try to coordinate the final steps to be able to run the jobs on
17.-19. April.

The packages will have to be finalized very soon, installed and testd on the sites, and we have

to decide on the macros to run. As a first step I would like to grasp the current status. Could you please help with this?

* Packages:
Can you Soeren, please summarize the status of the software. What is the current status, what is left to do?

* Macros:
The latest proposal by Soeren was (or did I miss something?)

```
-------------------------------------------
stt
===
root -b runsim.C"(nEvents,pT)"
root -b rundigi.C
root -b runreco.C

pT=30,40,50,...100 MeV/c
pT=100,200,300,...,1000 MeV/c
pT=1,2,3,...,7.5 GeV/c
10,000,000 events each

tpc
===
root -b run_sim_tpcmvd.C"(nEvents,pT)"
root -b run_rectrack_tpcmvd.C

pT=30,40,50,...100 MeV/c
pT=100,200,300,...,1000 MeV/c
pT=1,2,3,...,7.5 GeV/c
10,000,000 events each

dpm
===
Events and beam momenta for dpm will be proposed tomorrow
-------------------------------------------
```

Soeren, do you want to update?

What about the UrQMD and the fast sim macros Johan has already installed and tested on the grid?

Could you, Soeren and Johan, please try to make a concise list of possible simulations to run, based on the packages which will be available for the DC01?

Can you Johan please summarize the status of the macros and the jdls. Examples for UrQMD and the fast sim are obviously ready. What about the other simulations (stt, tps, dpm)? Will you be able to take care of this?

* Storage:
One event is 250 Bytes (the number was given by Soeren). Is this true for any of the prposed simulations (UrQMD?)? If yes then we most probably have far enough storage capacity (1TB ~ 4.E9 events).

* CPU time
Can somone give a (rough) estimate of the CPU time used per event? This number would help to decide on the number of events to submit (per job, total).

Cheers,

Paul

Subject: installation of relevant packages
Posted by Johan Messchendorp on Thu, 10 Apr 2008 15:03:11 GMT
View Forum Message <> Reply to Message

Hi all,

Once we have the most recent external packages installed, I will upload the most recent pandaroot revision including the macros which are proposed for the DC. This I can do during the weekend. I will also try to initiate a remote installation to all the other sites and report back where things went wrong (using the template on the wiki site).

Kind wishes,

Johan.

Subject: Re: one more week left
Posted by Jens Sören Lange   on Thu, 10 Apr 2008 16:32:01 GMT
View Forum Message <> Reply to Message

Hi Paul and all,

Quote:* Packages:
> Can you Soeren, please summarize the status of the software.
> What is the current status, what is left to do?

I can only comment on PandaRoot itself.

There were only 2 issues left.

1.) if riemannfit is on, then lhetrack cannot run.
This is more complicated and will need time to solve.
However, lhetrack is global (has tpc and mvd),
riemannfit is local (tpc only) in our understanding
So we decided to switch riemannfit off for the DC
(I am sorry to Sebastian and Tobias for this,

but we just have to make a decision just due to time pressure).

2.) include mvd into the tpc macros
(and so run the _global_ tracking for both mvd and tpc).
This was solved today by Stefano with help of Tobias and Ralf.
He checked it in.

-> this means:

rev---- is final for tpc, stt, dpm.

As the UrQMD code is not in the svn repository, this means:

rev2480 can be regarded as the PandaRoot version for the DC
(note:revised by Soeren on Friday Apr 11 15:37)

(only exception: if there is a last minute bug fix)

THIS MEANS:

(please read and enjoy)

we will run global(!) tracking for tpc digis(!) and mvd digis(!)

*Quote: Macros:
The latest proposal by Soeren was (or did I miss something?)
[...]
Soeren, do you want to update?

Yes, I would like to propose the beam momenta for the dpm simulations.
(these are input parameters for the bash scripts prepared by Johan).

[p_beam / GeV/c][what for?]

0.739   (for PhiPhi at 2.000 GeV)
2.202   (for PhiPhi at 2.500 GeV)
4.064   (for J/Psi 3.096 GeV)
6.234   (for Psi' 3.686 GeV)
6.571   (for Psi 3.770 GeV)
6.991   (for X 3.872 GeV)
7.277   (for Y 3.940 GeV)
7.705   (for Psi 4.040 GeV)
8.685   (for Y 4.260 GeV)
11.917   (for D*_sJ D_s 4.9178 GeV)
15.000   (for Drell-Yan)

I propose 10,000,000 events for each.

Quote:Could you, Soeren and Johan, please try to make a concise list of possible simulations
to run, based on the packages which will be available for the DC01?

Paul, can you specify your question a bit?

Because all the mentioned above is the list.

1. tpc sim and reco (see macros, nEvents, pT above)
2. stt sim and reco (see macros, nEvents, pT above)
3. dpm event generation
   (bash scripts by Johan, see p_beam, nEvents just a few lines above)
4. UrQMD event generation
   (see separate posting by Johan)

This is it.

Quote:* Storage:
One event is 250 Bytes (the number was given by Soeren). Is this true for any of the prposed simulations (UrQMD?)? If yes then we most probably have far enough storage capacity (1TB ~ 4.E9 events).

Paul, please be careful!!!

for stt and tpc sim and reco, it is much more!!!

It is the numbers in my last posting.

for tpc 56.5 kB=kiloBytes per event

for stt 17.3 kB=kiloBytes per event

plus an offset for each run (the "zero" events case).

cheers,

Soeren

Subject: Re: one more week left
Posted by Paul Buehler on Fri, 11 Apr 2008 07:50:45 GMT
View Forum Message <> Reply to Message

Many thanks for the update!!

Paul

Subject: if you can wait, please wait a few hours ...
Posted by Jens Sören Lange   on Fri, 11 Apr 2008 10:05:02 GMT
View Forum Message <> Reply to Message

Hi everyone,

just to let everyone know ...

I know we fixed the PandaRoot revision, but if there is a chance to wait for a few hours, please do so ...

Mohammad and Stefano are working on some more things.

The idea is to switch on a few more detectors (incl. EMC reco -> clusters for later track matching for both tpc and stt) and Mohammad is just writing a interface to call dpm from a root macro (so that the bash scripts are not needed). We also discussed to run dpm full sim (Geant3, Geant4) instead of only dpm event generation.

So, -> a few nice things!

Anyway, if there is any chance that you wait for a few hours, this would be great ...

cheers, Soeren

Subject: Re: if you can wait, please wait a few hours ...
Posted by Florian Uhlig on Fri, 11 Apr 2008 12:12:16 GMT
View Forum Message <> Reply to Message

Hi everyone

I installed the latest version of the external packages which is called panda_extern::apr08. This is the version from march with some small changes for the grid. Installation at GSI works fine, so i didn't expect problems with the other sites.

I will prepare everything for the installation of pandaroot.

Ciao

Florian

Subject: final PandaRoot for DC
Posted by Jens Sören Lange   on Fri, 11 Apr 2008 13:50:57 GMT
View Forum Message <> Reply to Message

Hi all,

we proudly present

PandaRoot rev2480
-> this will now be the version for the DC.

/macro/dc1/stt

new: contains emc digis (clusters), too

for track matching

note: there is no stt rundigi.C macro anymore
-> it is not integrated into the runreco.C macro
(as it is also for tpc)

/macro/dc1/tpc

new: contains emc digis (clusters), too
for track matching

track visualization is now switched off
-> saves disk space

/macro/dc1/full

THIS IS NOW THE DPM !

(it means, basically no basf script necessary anymore,
but if you want to run the DPMGen executable for some
reason, this also still works).

new: we do not only event generation but even Geant3.

all detectors.

pT and nEvents are now parameters of the macro.

disk space: 105 kB per 1 event

(so not anymore only 250 bytes/event!)

CPU time: ~1 sec per 1 event.

note: there is no option for the seed.
Mohammad says that it is taken from the date.

NOTE! pgenerators/DpmEvtGen needs to be compiled on each site!
(this is not part of the cbuild)
but Florian will prepare it for the packages.

all is tested and works.

additional note:

all macros contain new newest mvd and newest forward emc geometry!

(they were generated with different external packages,
but Stefano fixed it).

Thanks everyone for debugging etc. (Mohammad, Stefano, Ola, Tobias, Florian, ...)

I think the phone companies will be very happy about today's accounting bill.

Subject: random number generator for DPMgen
Posted by Johan Messchendorp on Fri, 11 Apr 2008 15:06:21 GMT
View Forum Message <> Reply to Message

Hi,

One very important issue. How is the random seed for the new DPM interface set?

Will the instruction

gRandom->SetSeed(<my favorite seednumber>);

in the macro do the job for me?

Johan.

Subject: Re: Upcoming DC - PandaRoot rev2480
Posted by Dan Protopopescu on Fri, 11 Apr 2008 17:01:22 GMT
View Forum Message <> Reply to Message

Also from me lots of thanks to Soeren and everyone on Soeren's list. Great job!

When everything is final, could we have a list with the package names to be installed for the DC posted on the wiki (as reference for the site admins) ?
http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1#Packages

Have a nice weekend.

Subject: Re: random number generator for DPMgen
Posted by Mohammad Al-Turany on Fri, 11 Apr 2008 17:44:43 GMT
View Forum Message <> Reply to Message

Hallo Johan,

Quote:Will the instruction
gRandom->SetSeed(<my favorite seednumber>);
in the macro do the job for me?

Not really, because dpm do not know anything about root and gRandom.

in the PndDpmDirect, the seed is set using:

```
Long_t Time = time(NULL);
int a = Time/100000;
seed = Time - a*100000 + a/100000.;
```

which is the same as it was in the executable, except that it is changed for each event, so
each call to the next event will set the seed again. It is simple but I think it is enough, with the
executable you set the seed once for all events and here you change it for each event! Any
way if you think this is not enough we can change it and take the seed for example from
gRandom, this should be straight forward!

regards

Mohammad

Subject: Re: random number generator for DPMgen
Posted by Johan Messchendorp on Fri, 11 Apr 2008 20:20:50 GMT

Hi Mohammad and all others,

Thanks for your clear answer, Mohammad. I just wonder what the probability is that two (or
more) jobs will start a DPM event with the same seed. The "time(NULL)" instruction gives an
output in precision of seconds, right? Hmmm, in that case it could actually be very likely. It
might, therefore, be more safe to set it up it via gRandom, which one can more easily control
from the "outside". Actually, is this the same for the box generator (as used for the other DC
macros)?

Best wishes,

Johan

(ps, really nice work what has been done for this DC. Thanks to everyone involved, really
impressive!!!!!)

Subject: Re: random number generator for DPMgen
Posted by Mohammad Al-Turany on Sat, 12 Apr 2008 17:46:30 GMT

Hi,

The box generator uses gRandom, the code in PndDpmDirect was taken from the original
DPM files! Anyway I change it now (rev. 2487) to use gRandom for setting the seed.

regards

Mohammad

Subject: Re: random number generator for DPMgen
Posted by Johan Messchendorp on Sat, 12 Apr 2008 17:58:05 GMT

Thanks Mohammad!

I also added a seed input to all the DC1 macros, and set gRandom accordingly to the seed input. This modification has been included in rev.2487. as well.

Johan.

Subject: Re: Upcoming DC - Testjobs to be specified and documented
Posted by Kilian Schwarz on Tue, 15 Apr 2008 08:41:19 GMT

Hi all,

would it be possible to put, preferably to the Wiki, a list of testjobs which are to be submitted by the site admins for various purposes ?
So far only the urqmd testjob is well documented. About everything else the information is not so clear for my feeling.

Cheers,

Kilian

Subject: Re: Upcoming DC - Testjobs to be specified and documented
Posted by Paul Buehler on Tue, 15 Apr 2008 08:45:33 GMT

ok - I'll add some instructions for running test jobs.

Paul

Subject: Re: Upcoming DC - Testjobs to be specified and documented
Posted by Paul Buehler on Tue, 15 Apr 2008 09:23:13 GMT

Dear all,

I added a section 'Running test jobs' on the wiki.

Paul

Subject: Run list for DC1 open for discussion
Posted by Paul Buehler on Wed, 16 Apr 2008 14:51:05 GMT
View Forum Message <> Reply to Message

Dear all,

I put a list of jobs to run during the DC on the wiki
http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1 and would like to ask you for
your comments and suggestions.

The list contains all the proposed scripts and momenta. However it foresees to produce only
100'000 events per momentum and not the 10'000'000 as originally proposed.

There are:

- 240 stt splits a 10'000 events, each running for ~2h, producing 200 GB in total
- 240 tpc splits, 10'000, ~2h, ~200 GB
- 110 full splits, 10'000, ~9h, ~200 GB
- 100 urqmd splits, 10'000, ~1h, ~1 GB
- 1000 generic splits, -, ~10 sec, 10 MB

In total: ~2000h, 1 TB

The critical point I think will be the available CPU time and data transfer time.
Alternatively we could run only a few momenta with higher statistics.

Paul

Subject: Re: Run list and transfer bandwidth
Posted by Kilian Schwarz on Wed, 16 Apr 2008 14:55:36 GMT
View Forum Message <> Reply to Message

Hi Paul,

great !!!
Can we make a guess concerning the needed bandwidth for inter site transfers ?
I mean, we have a certain amount of promised CPUs at a site.
This results in a certain amount of jobs and finally in a certain amount of produced output per
time unit which needs to be transferred to certain places within a given time which results in a
need bandwidth of X Mbs between sites.

Cheers,

Kilian

Subject: Re: Run list and transfer bandwidth

Posted by Dan Protopopescu on Wed, 16 Apr 2008 14:59:08 GMT
View Forum Message <> Reply to Message

I propose Kilian calculates that.

Subject: Re: Run list and transfer bandwidth
Posted by Kilian Schwarz on Wed, 16 Apr 2008 15:16:13 GMT
View Forum Message <> Reply to Message

and I thought Dan would do that ...

Cheers,

Kilian

Subject: Start of DC
Posted by Paul Buehler on Wed, 16 Apr 2008 15:18:39 GMT
View Forum Message <> Reply to Message

Dear all,

This is to inform you, that I plan to start the job submission for the DC, on April 17, at 08:00 AM Vienna time.

Please do not interfere and submit jobs after that. I will announce the end of the DC, hopefully not later than on April 21.

I'll try to regularly update the wiki
http://panda-wiki.gsi.de/cgi-bin/viewauth/Computing/DataChallenge1 with some status reports to keep you informed.

Cheers,

Paul

Subject: Re: Start of DC - missing executables
Posted by Kilian Schwarz on Thu, 17 Apr 2008 04:40:24 GMT
View Forum Message <> Reply to Message

Hi Paul,

it seems as if the executables for the DC1 have not been registered correctly.
I get:
<307] /panda/user/p/pbarprod/bin/ > whereis run_stt_dc1.sh
Apr 17 06:38:47 info   The file bin/run_stt_dc1.sh is in
<307] /panda/user/p/pbarprod/bin/ > whereis run_tpc_dc1.sh
Apr 17 06:38:54 info   The file bin/run_tpc_dc1.sh is in

<307] /panda/user/p/pbarprod/bin/ > whereis run_urqmd.sh
Apr 17 06:39:01 info   The file bin/run_urqmd.sh is in

but for example:
<307] /panda/user/p/pbarprod/bin/ > whereis sim_ana_chain.sh
Apr 17 06:39:44 info   The file bin/sim_ana_chain.sh is in
          SE => PANDA::Glasgow::raid0 pfn =>
root://panda.gla.ac.uk:8222//raid/SEData/02/20001/f84537b8-733f-11dc-865
c-00e081408d5a.1191587949

which is the correct output, I believe.
We should have a storage element assigned to the files.

Cheers,

Kilian

Subject: Re: Start of DC - missing executables
Posted by Paul Buehler on Thu, 17 Apr 2008 05:40:56 GMT
View Forum Message <> Reply to Message

Indeed - the executables seem to have disappeared.

Might be, that this is connected with the fact that Glasgow is currently down?

Paul

Subject: Re: Start of DC - missing executables
Posted by Kilian Schwarz on Thu, 17 Apr 2008 05:48:16 GMT
View Forum Message <> Reply to Message

Hi Paul,

but PANDA::Glasgow::raid0

seems to be online

I can also register files to Glasgow.

Can you reregister them ?
e.g. at GSI ?

Please use:
PANDA::GSI::virtual

for the DC.

NOT
Panda::GSI::file

Cheers,

Kilian

Subject: new SE at GSI
Posted by Kilian Schwarz on Thu, 17 Apr 2008 05:49:48 GMT
View Forum Message <> Reply to Message

Dear all,

for the upcoming data challenge:

please use

panda::gsi::virtual

and not

panda::gsi::file

the new SE is an xrootd cluster with significant more storage space than the old NFS mounted disk.

But both SEs are still operational, in case you want to access old files.

Cheers,

Kilian

Subject: Re: Start of DC - missing executables
Posted by Paul Buehler on Thu, 17 Apr 2008 06:11:24 GMT
View Forum Message <> Reply to Message

I did re-register the scripts. They are now at PANDA::Vienna::file.

Paul

Subject: Re: Start of DC - missing executables
Posted by Kilian Schwarz on Thu, 17 Apr 2008 06:19:08 GMT
View Forum Message <> Reply to Message

thanks

Cheers,

Kilian

Subject: DC1 finished
Posted by Paul Buehler on Mon, 21 Apr 2008 06:27:26 GMT

Dear colleagues,

The first PANDA grid Data Challenge, which was running since last Thursday morning is now finished. A brief summary of what has been happening in the past few days can be found at http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1#Results.

This DC was very helpful. We will have to analyze and evaluate the result. But I am sure, that the experience we gained will help to further improve the procedures, the system, and its operation.

Many thanks to all of you who have helped in the last days to get and keep the system running.

Best regards,

Paul

Subject: Re: DC1 finished
Posted by Kilian Schwarz on Mon, 21 Apr 2008 11:23:05 GMT

Hi Paul,

thanks a lot for the great work you did.
Why did you put the RUNNING distribution in the Wiki ?
If you put the DONE distribution it looks quite different. According to the DONE of last week in ML GSI computed about 50% of the jobs, while the other sites shared the other 50%. Ateneo contributed with almost 6%, which is great with 10 machines.
What was the main reason why the STT jobs did not work ?
Also in the full_dc quite some jobs seem to have failed. What was the main reason here ?
All in all we seem to have close to 50% error rate, which needs to be improved, somehow.

Cheers,

Kilian

Subject: Re: DC1 finished
Posted by Dan Protopopescu on Mon, 21 Apr 2008 11:33:03 GMT

Dear all,

We are still updating the DC1 wiki. I will add final plots and graphs from MonaLisa shortly.
Please watch this for updates.

Subject: DC1 Summary
Posted by Dan Protopopescu on Mon, 21 Apr 2008 14:21:43 GMT
View Forum Message <> Reply to Message

Dear all,
I updated the wiki with more comments and plots. Please review the updated info at:
 http://panda-wiki.gsi.de/cgi-bin/view/Computing/DataChallenge1#Evaluatio n
We will update the summary of the physics data produced asap.

Subject: Re: DC1 Summary
Posted by Kilian Schwarz on Tue, 22 Apr 2008 06:09:45 GMT
View Forum Message <> Reply to Message

great plots and description. Very informative.
I think the whole exercise was a big success !!!

Cheers,

Kilian