Structuring ZEUS Common ntuples; Gutcode modification for dCache

Summerstudent Program 2007, DESY

Oleksandr Volynets

National Taras Shevchenko University of Kiev

14 September 2007

Supervisor: Achim Geiser

Abstract

At present report my work with ZEUS Common ntuples and dCache is described. It is applied to the so-called Gutcode analysis framework. Also a .root script was developed to generate mini_ntuples from Common ntuples. It is useful when only a subset of the variables from the ntuples is needed and it significally saves time.

Contents

1 Introduction

1.1 Overview

Several different directions of activity in the framework of the ZEUS experiment are described. They all are connected with dCache and Common ntuples [1] that are now in development. Soon dCache should be used in order to save disk space, such that every root-based analysis software can read data directly from tapes. The corresponding data format is called "Common ntuples". Soon it will be stored on tapes and it will needed to be readable by any analysis software.

1.2 Mini ntuples files idea

The suggestion of common ntuples was developed to keep all data from ther ZEUS detector at HERA in root and paw files for the best perfomance such that every group can work with the same data. These files include information about many parameters of tracks of the particles, their vertices etc. It consists of more than 600 parameters. Usually for analysis only some of them are needed (20-50). In this report the idea of structuring common ntuples files by root script is described. This script can convert Common ntuples files to new Mini_ntuples which consist only the needed parameters for further analysis. A user interface consist of changing input, output files, changing list of the parameters one needs and some more.

1.3 Gutcode and dCache

There is a Gutcode [3] framework used by a group of people from the ZEUS heavy flavour group for various analyse. It was written under Root version 5.08.00. As the idea of Common ntuples is that every software can read data for the analysis directly from dCache, Gutcode is also should provide this possibility. There is also another software ZACON [2] (by the OFFLINE group) which provide reading the data from dCache. But it needs Root version 5.14.00, and original Gutcode from the CVS repository crashes under Root 5.14.00. In the present report modifications for Gutcode to fix this are described.

2 Working with Common ntuples

Now Common ntuples files are still in development and until the 2010 all data from detectors will be kept in this format and will be stored to dCache to save place on local machines (zeniths). To use data from Common ntuples from dCache there are 2 possibilities:

- To do the analysis directly from dCache. This is very time consuming. Every time the analysis needs to be redone and it will take much time.
- To make Mini_ntuples with only the needed parameters for the analysis and then work with the trelatively small files. This is much faster.

There is a script written by me, in which only the necessary branches from the Common ntuples files are written into new Mini_ntuples files. It is a simple program written in Root. You can download it and manual on it from
`http://www-zeus.desy.de/~volynets/Mini_ntuples/`
There are also include files, which can be changed according to necessity:

- *./inculde/Input.inc* where input files are enumerated (*data_05e_57002*.root ...*);
- *./include/Blocks.inc* where input orange blocks are enumerated (*Tracking.inc, V0lite.inc ...*);
- *./include/Tracking.inc* etc., where parameters according to there orange blocks are enumerated (*trk_ntracks, trk_type ...*).

Also some optional changes may be done in the script like paths to input and output files etc. Everything is described in the manual on a web page [4]. The conversion made by this script is shown on Fig. 1.

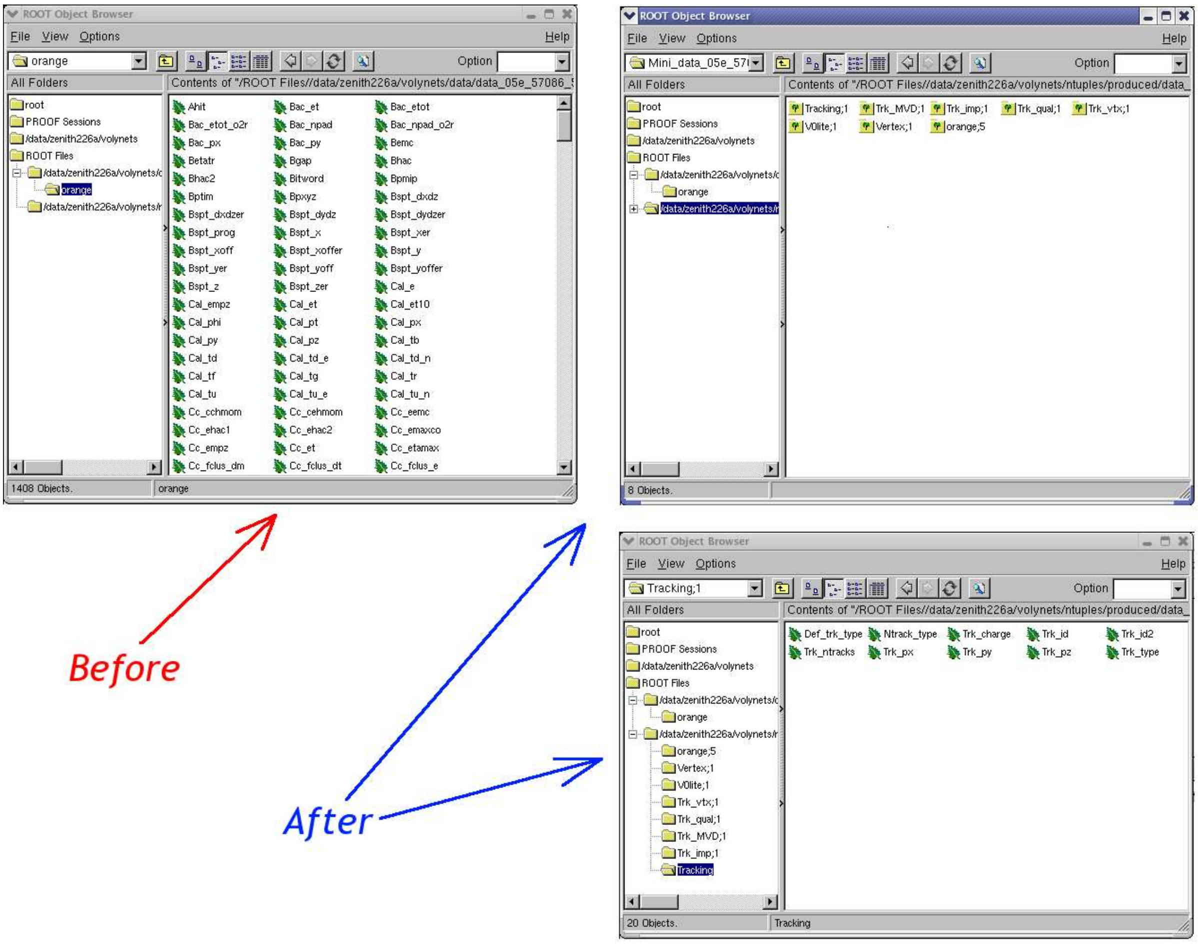


Figure 1: *Illustration of script effect.*

This conversion can be done in two ways:

- To separate blocks according to their ORANGE block (shown in Fig.1).
- To keep all parameters together as it is in initial Common ntuples (the same tree "orange").

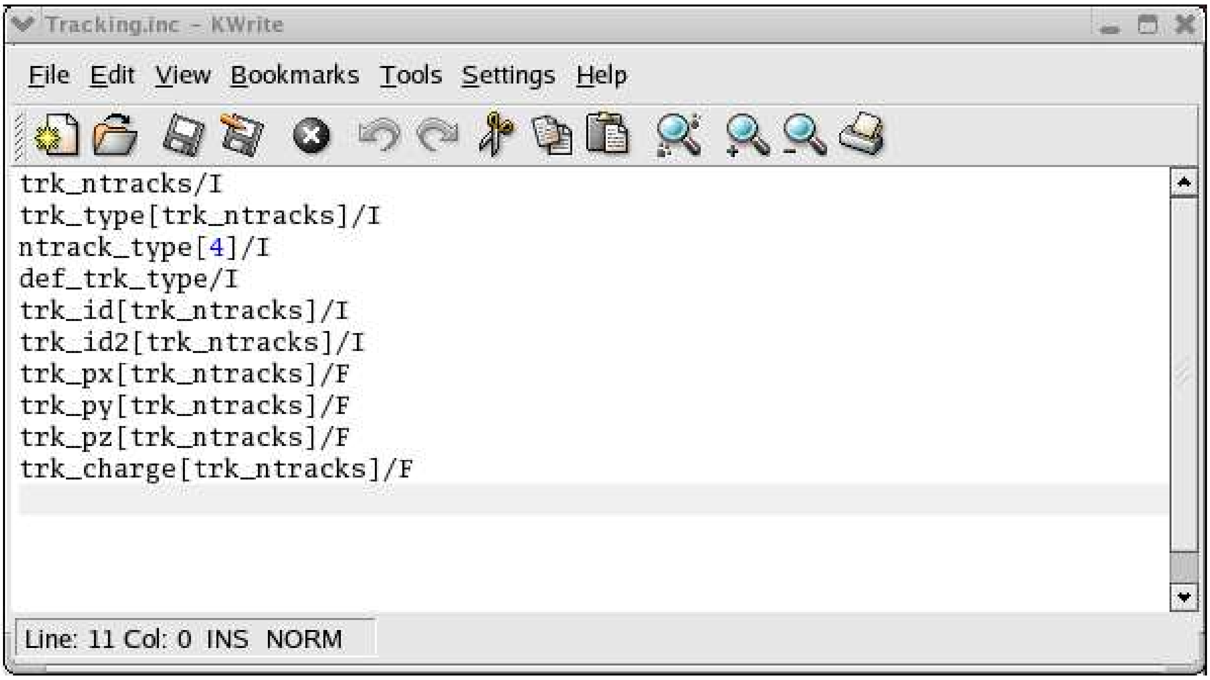
It is controlled by a flag in script and described in manual [4].

Advantages of this script are:

- It is not a source code, it doesn't need to be compiled etc. This is .cxx (root) script.
- Saves much time to further analysis after mini_ntuples had been generated.
- Users can change what parameters of each block they need and don't need (Fig. 2).

Plans of developing the script:

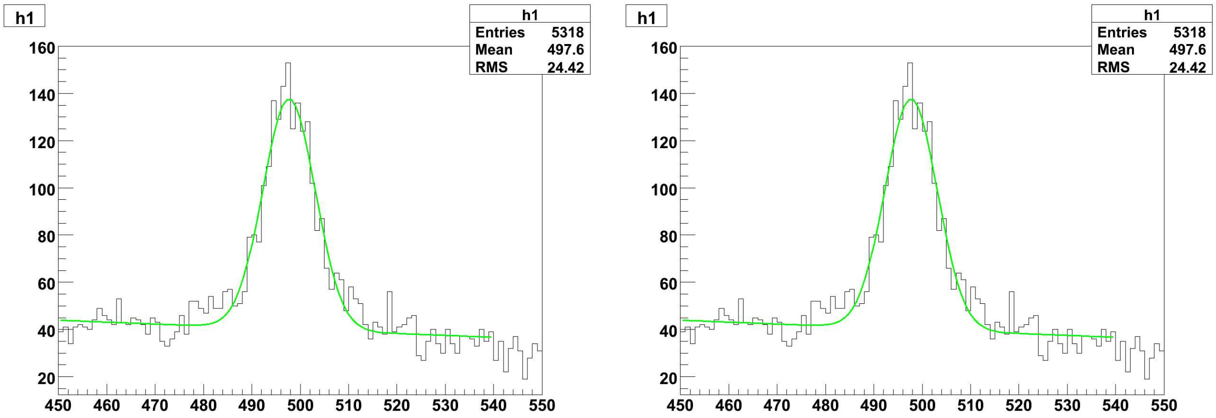
- To add other include files with appropriate blocks, because now it consists of only several blocks for demonstration.



```
trk_ntracks/I
trk_type[trk_ntracks]/I
ntrack_type[4]/I
def_trk_type/I
trk_id[trk_ntracks]/I
trk_id2[trk_ntracks]/I
trk_px[trk_ntracks]/F
trk_py[trk_ntracks]/F
trk_pz[trk_ntracks]/F
trk_charge[trk_ntracks]/F
```

Figure 2: *You can delete branches you don't need.*

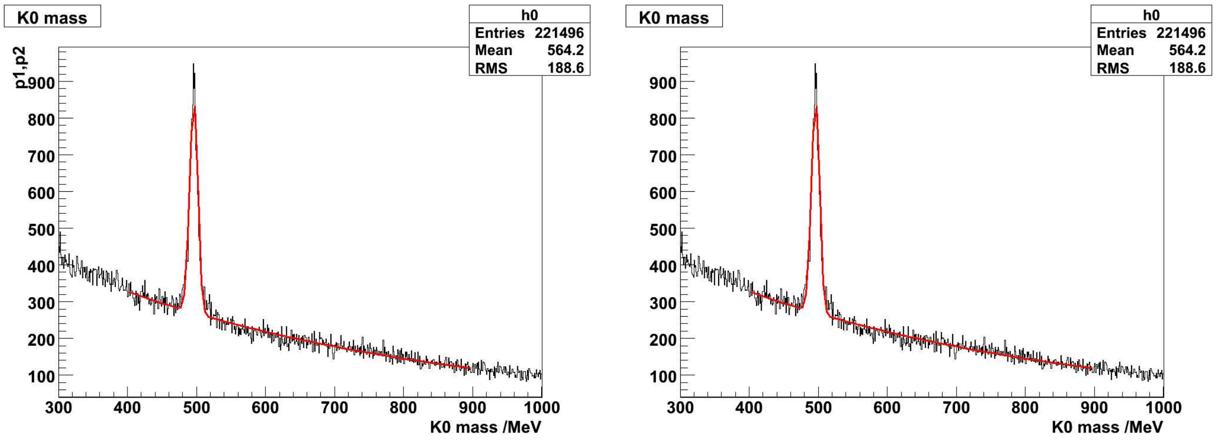
Some examples of distributions was made after Mini_ntuples had been generated and it's comparison with initial histograms from Common_ntuples is shown in Fig. 3, 4 and 5. They are the same as it should be.



(a) From ntuples

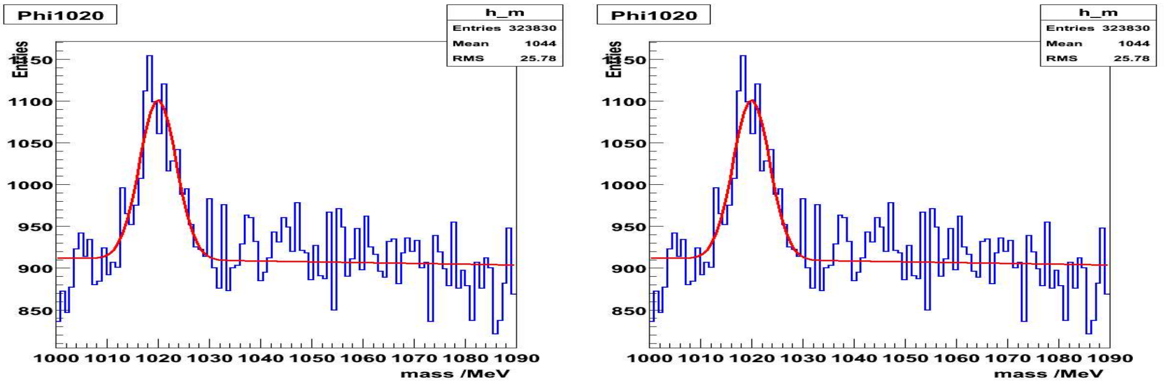
(b) From mini_ntuples

Figure 3: $K^0 \rightarrow \pi^+ + \pi^-$.*Blocks: Tracking, Trk_vtx, Trk_MVD, Trk_imp, Trk_qual, Vertex, Trk_vert, V0lite.*



(a) From ntuples (b) From mini_ntuples

Figure 4: $K^0 \rightarrow \pi^+ + \pi^-$. *Blocks: V0lite*



(a) From ntuples (b) From mini_ntuples

Figure 5: $\phi \rightarrow K^+ + K^-$. *Blocks: Tracking, Trk_vtx, Trk_MVD, Trk_imp, Trk_qual, Vertex, Trk_vert*

So the *conclusion* is that all Mini_ntuples files were written in the right way.

3 Gutcode and dCache

3.1 Problems working with dCache

As all data is going to be kept in dCache, every framework for analysis should provide possibility to use dCache also. It can be done in 3 ways:

- Include all analysis features to ZACON. **Problems:** It will need much time and to do it in such way, every analysis tool needs to be included to ZACON.
- Include ZACON tools to Gutcode. **Problems:** Almost the same: ZACON code should be included in every analysis tool.
- Use "zeus://" protocol, provided by ZACON for common use. **YES!!** So now the data from dCache can be read in the same way as from local machines(zeniths).

The problem for Gutcode in using ZACON is that it can work only with Root 5.14.00, but Gutcode from the CVS repository crashes with Root 5.14.00. Some modifications of Gutcode were made for compatibility with Root 5.14.00. How to work with the ”zeus://“ protocol is described below.

3.2 Modifications for Gutcode

The modifications are:

- The ”Append“ function should be excluded when writing a small_ntuple in GAnalysis.cxx: 407. Writing eventlist to file should be done in such way:

```
if( gCards->GetDoEventlist() )
{
  ffile_ev->Write();
  ffile_ev->Close();
  ffile_ev = 0;
}
```

- Macros ”ForEach“ should be replaced by ”R__FOR_EACH“ in GAnalysis.cxx: 364, 367, 376, 403.

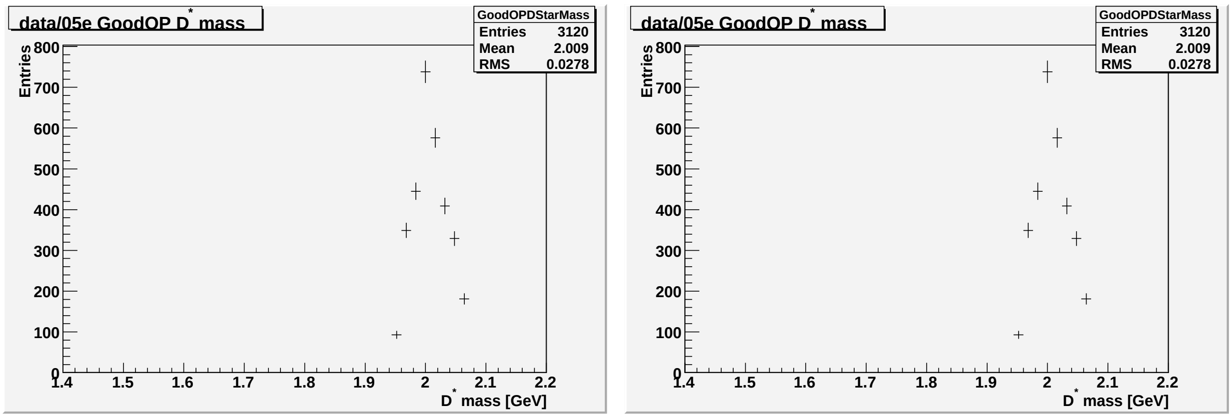
If you want to read data directly from dCache you should add files

$GUTCODEDIR/analysis/dcache/ABTdata.txt,

where AB is the year of data, T is type of data (e/p - electron, positron). These files should contain the full path to dCache data:

zeus://acs/z/ntup/05/v01/root/data_05e_57014_57014_02.root

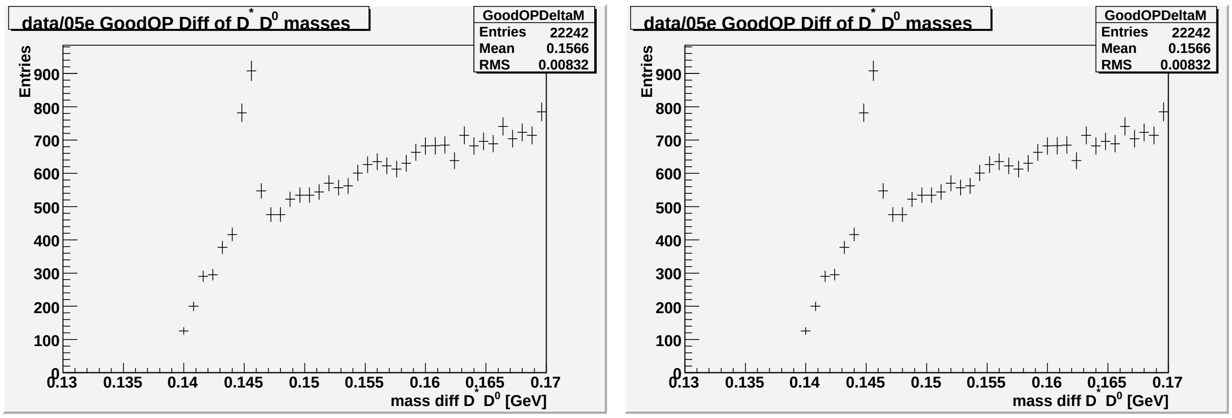
Here are some histograms generated with Gutcode. There was used Dmitriy Gladkov's setup for D^* and D^*-muon analysis etc. One of pictures was generated when reading from zenith, another when reading from dCache for 2005e data:



(a) From zenith

(b) From dCache

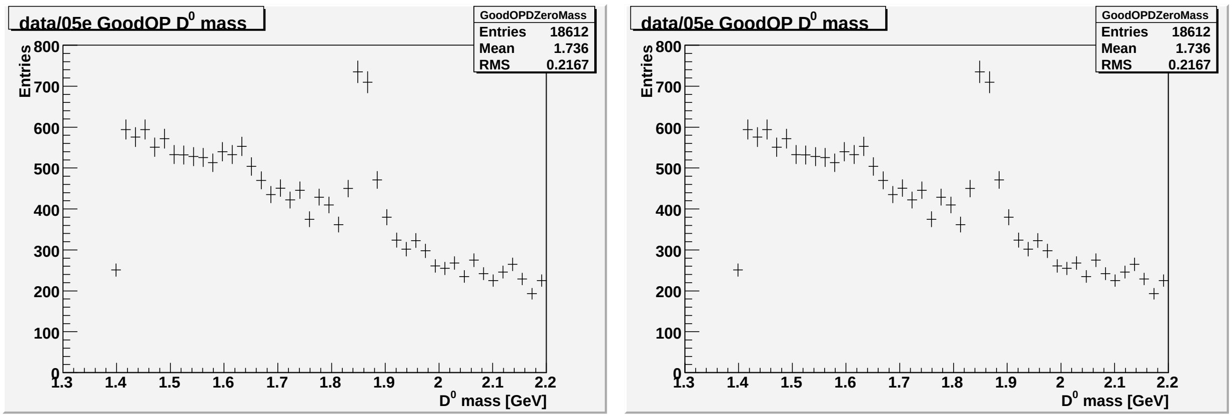
Figure 6: D^* *mass*



(a) From zenith

(b) From dCache

Figure 7: *Mass difference:* $D^* - D^0$



(a) From zenith

(b) From dCache

Figure 8: *D0 mass*

Also there are another pictures of distributions from HERA I data (1996-2000) using the same Gutcode (Fig. 9, 10):

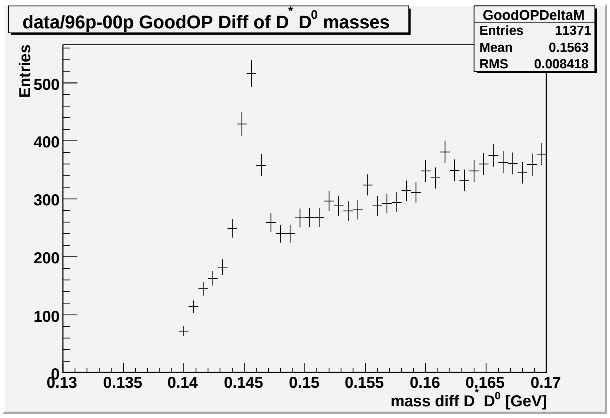


Figure 9: *Mass difference:* $D^* - D^0$

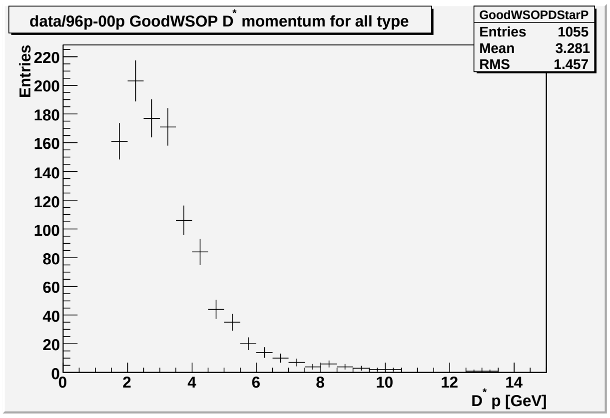


Figure 10: D^* *momentum.*

So the *conclusion* is that Gutcode is backward compatible with old data and it can be used for all data analysis.

3.3 Conclusions for another analysis frameworks

Any another analysis framework can have direct access to dCache as provide by ZACON. You need to migrate to Root 5.14.00. Two libraries should be implemented before reading files in your source code:

- gSystem->Load("libdcap.so");
- gSystem->Load("libzio.so");

Then the procedure of reading files is very simple:

- TChain *fChain = new TChain(”orange”);

- fChain -> Add ("zeus://acs/z/ntup/05/v01/root/*.root");

To increase the speed of reading from dCache the cache should be implemented:

- export DCACHE_RAHEAD = 1
- export DCACHE_RA_BUFFER = {size in bytes}
  recommended value is 100000

So after that you can write
zeus://acs/z/ntup/05/v01/root/*.root
when you read your .root data files instead of
/data/zenith224c/gmuon/RUNS/data_05e_v2007a/data/*.root

4 Conclusions

4.1 Common ntuples

Now that the data format is migrating to Common ntuples format. So some utilities that provide flexible using it may be useful. My idea was to struct common ntuples files and produce then mini_ntuples files. It provides rather flexible and simple mechanism of generating small files that then can be used for different analysis. This idea was realized in ntuples_to_mini_ntuples.cxx utility.

4.2 dCache tools

As the data format is migrating to Common ntuples format, it is also migrating to tapes, so reading data from dCache is a necessity of present and further work. The problem is to make every software to be able to use dCache. As an example Gutcode was examined. It was managed to convert this software to Root 5.14.00 to be able to use tapes-reading tools. So now it can easily work with dCache as such as with local machines.

5 Future works and plans

In the future I am going to support my program with future modifications of data format etc. and it will be very nice of getting some reviews of everybody is interesting in it. Also as I was working with Gutcode I probably will be able to support this tool.
Also I am going to do my diploma in physics and I plan to do my PhD also here, in ZEUS collaboration.

6 Acknowledgements

I would like to thank my supervisor Achim Geiser for coordinating everything I was doing. Thanks also to Dmitriy Gladkov for giving his Gutcode setup to me, Igor Rubinskiy, Benjamin Kahle, Philipp Roloff, Ingrid-Maria Gregor for discussion different problems, and Tobias Haas, Elisabetta Gallo, Joachim Meyer, all ZEUS collaboration and DESY for the opportunity to participate in Summer Student program. Thanks also to Vladimir Aushev for giving many advices during my being here.

References

[1] Common ntuples homepage:
`http://www-zeus.desy.de/ZEUS_ONLY/analysis/comntp/`

[2] ZACON homepage:
`http://www-zeus.desy.de/ZEUS_ONLY/zacon/twiki/bin/view`

[3] Gutcode homepage:
`http://www-zeus.desy.de/~gutsche`

[4] Ntuples_to_minintuples link:
`http://www-zeus.desy.de/~volynets/Mini_ntuples`