HERA-DP@Max-Planck Institut für Physik

Andrii Verbytskyi

DPHEP Collaboration Workshop
Geneva,
June 5, 2015

Content

- **HERA-DP@MPP status**
- **Technical details**

HERA-DP@MPP status

Introduction and goals

The MPP aims to preserve the data of the following experiments:

- **H1@HERA**
- **ZEUS@HERA**
- OPAL@LEP
- JADE@PETRA

The main intention is to provide facilities for the physics analysis and to do physics analysis for in house experiments.

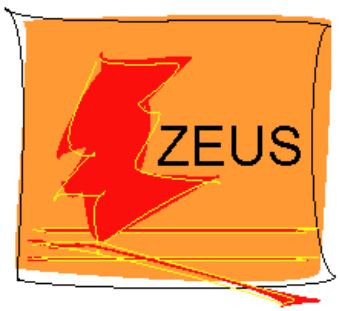
HERA-DP@MPP

The general strategy is:

- Save the bits(files) $\rightarrow$ copy the data from DESY to RZG[1];
- Provide an access to the saved bits(files) $\rightarrow$ define access and archiving policy;
- Save the software $\rightarrow$ provide installation of virtual machines with the software;
- Save the documentation $\rightarrow$ rely on DESY and open databases like InSpire;
- Share experience $\rightarrow$ provide documentation for the software preservation and software deployment.

[1] Rechenzentrum Garching, Gießenbachstraße 2, D-85748 Garching

ZEUS-DP@MPP



Bits for ZEUS@MPP

ZEUS data is copied to RZG:

- FIXME Tb
- 1,2 Mio files

Access for ZEUS@MPP

- ZEUS data is accessible worldwide via gsidcap, webdav, gridftp, srm protocols:
  - https://grid-dav.rzg.mpg.de:2880/zeus/
  - srm://grid-srm.rzg.mpg.de:8443/pnfs/rzg.mpg.de/data/zeus/
  - gsidcap://grid-srm.rzg.mpg.de:22128/pnfs/rzg.mpg.de/data/zeus/
  - gsiftp://grid-gftp2.rzg.mpg.de:2811/pnfs/rzg.mpg.de/data/zeus/
- To access the data via all protocols grid certificate and ZEUS Grid VO membership is needed;
- The tape back-end for RZG is in preparation. The plan is to have data and most "popular" MC samples always on-line on disk pools and the rest of the data on the tape.

Documentation for ZEUS@MPP

- MPP policy is to rely on DESY and open databases like InSpire.
  - http://zeusdp.desy.de
  - http://inspirehep.net/search?p=collaboration:%27ZEUS%27&ln=en

Software preservation for ZEUS

- The software is preserved in a form of CentOS5/CentOS6/CentOS7 x86_64 customised ISO images.
- The images include ZEUS software packaged in default RedHat rpm packages.
- The images can be installed automatically and w/o internet access via default build-in RedHat anaconda installer.
- The images suitable for real machines and all kind of i386/x86_64 emulators with CD drive support.
- The images can be found on https://wwwzeus.mpp.mpg.de/dphep.html (use default zeus password or contact me).
- Please, download them and test.

Documentation for ZEUS-DP@MPP

- A short documentation is available on https://wwwzeus.mpp.mpg.de/dphep.html.
- A set of scripts for creation of customised ISOs is provided as well.
- Example-based, future-oriented documentation is the best approach. The requests and answers to some questions should be added to the web-page.

Example of ZEUS data usage: $D^{*\pm}$ mass peak

```
#include <string>
#include <vector>
#ifdef __CINT__
gSystem->Load("/usr/lib64/libCNINFO.so");
int main()
{
#else
#include "CNINFO.h"
int main(int argc, char **argv)
{
     TApplication* theApp = new TApplication  ("tapp", &argc, argv);
#endif
     CNINFO *A= new CNINFO();
     std::vector<std::string> test=A->GetFilesList("v08b","07p");
     TChain* C= new TChain("orange");
     for (int i=0; i<5; i++)
          {
               printf("%i:%s\n",i,test.at(i).c_str());
               C->Add(test.at(i).c_str());
          }
     TCanvas* T= new TCanvas();
     T->cd();
     C->Draw("Ds1ms","Ds1ms>0.14&&Ds1ms<0.17","",1000,1000);
     T->SaveAs("Ds1ms.pdf");
#ifndef __CINT__
     theApp->Run();
#endif
     return 0;
}
```

ZEUS-DP@MPP TODO

- Provide more examples on analysis;
- Provide an example for MC generation and cninfo data base update;
- Make virtual machines run on cloud services, e.g. CERN OpenStack.

If you have an analysis to start you can contact me and I can try to help to set it up in DP mode.

H1-DP@MPP

Bits for H1@MPP

Only part H1 data is copied to RZG (the decision on the datasets to preserve was made only in May 2015):

- FIXME Tb
- FIXME Mio files

Anticipated end of copying: 08.06.2015.

Access for H1@MPP

- H1 data is accessible worldwide via gsidcap, webdav, gridftp, srm protocols:
  - https://grid-dav.rzg.mpg.de:2880/hone/
  - srm://grid-srm.rzg.mpg.de:8443/pnfs/rzg.mpg.de/data/hone/
  - gsidcap://grid-srm.rzg.mpg.de:22128/pnfs/rzg.mpg.de/data/hone/
  - gsiftp://grid-gftp2.rzg.mpg.de:2811/pnfs/rzg.mpg.de/data/hone/
- To access the data via all protocols grid certificate and H1 Grid VO membership is needed;
- The tape back-end for RZG is in preparation. The plan is to have data and most "popular" MC samples always on-line on disk pools and the rest of the data on the tape.

Documentation for H1@MPP

- MPP policy is to rely on DESY and open databases like InSpire.
  - http://FIXME
  - http://inspirehep.net/search?p=collaboration:%27H1%27&ln=en

Software preservation for H1

- The software will be preserved in a form of CentOS5/CentOS6/CentOS7 x86_64 customised ISO images.
- The images are not ready yet, but should be similar to those of ZEUS.
- As the H1 software is going to be recompiled the deployment will be more complicated than for the ZEUS software.
- ...
- The location of images will be specified.

Documentation for H1-DP@MPP

The H1 documentation for H1-DP@MPP will be close to H1-DP@MPP.

- A simple documentation will be published onine.
- A set of scripts for creation of customised ISOs will be provided as well.
- Example-based, future-oriented documentation is the best approach. The requests and answers to some questions should be added to the web-page.

Technical details

Software preservation

The offline software in many experiments can be splitted in four parts:

- Data analysis framework (e.g. ROOT/PAW based tools etc.);
- MC production package (set of generators, detector simulation with GEANT etc.);
- File catalogue (e.g. SQL databases);
- Event display.

Software preservation

The virtual machines are build from original CentOS ISO images with some customisation:

- The experiment-specific software is packed in rpm packages and put on the ISO image;
- Some other rpm with software (e.g. compilers of HEP libraries) is added;
- kickstart file for an automatic installation w/o internet access is added as well.

Software preservation problems

With the chosen model of data preservation the most complicated task is to deploy the software in an isolated environment of virtual machines.

- Hard-coded paths in software packages;
- Absence of configuration, compilation and deployment scripts;
- Absence of control version systems and dependency tracking between packages;
- Absence of source code(!);
- Strong dependence on environment: OS type, compilers, system libraries;
- Using non-standard self-made tools even if a well-established standard alternative exists.

ZEUS software preservation problems

ZEUS-specific software is splitted in four parts:

- Data analysis framework is ROOT5+PAW;
- MC production package (ZMCSP) is a set of generators, event files conversion utilities, detector simulation with GEANT3;
- File catalogue (CNINFO) is a text file with a parsing script and SQLite database with ROOT/C++ interface;
- Event display (ZEVIS) is Qt/ROOT-based;

In addition packages that are widely used in HEP should be available: g++, gfortran/g77, python, make, CLHEP, cernlib, fastjet, HEPMC, pythia8 and others.

ZEUS software

- The ROOT5 of the latest version and PAW are available for all platforms from default repositories as rpm packages.
- The MC standalone package, ZMCSP includes many precompiled executables and data files, so the total size is 11Gb. Unfortunately to reduce the size of the package a deep expertize is needed. The size of the package slows download the installation.
- The original file catalogue was located on a separate server with SQl database. A Perl-based utility cninfo was used to extract the information from the DB. DP@DESY planned to dump it to ASCII file and provide a shell script that mimic cninfo behaviour. Much faster and much more portable DP@MPP solution is to use SQLite3 database and ROOT/C++ class for the direct access to the filecatalog from the user analysis. Software for both options were packaged as rpm package.
- The ZEVIS event display was repackaged as rpm and has an option to be used outside of DESY.

What we can learn from software preservation problems

- Unfortunately the HEP software is very rarely packaged in any standard way. As usual there is no rpm or deb packages, packages don't have "configure" scripts, compiled "in place" and so on. Introducing as least some standard ways for the software installation is extremely helpful for the portability. Famous examples are fastjet or clhep.
- In many cases it is important to split the software into smaller pieces with well-defined dependence.
- Using well-established formats for data bases and other files can help to avoid problems.

Bit preservation problems

- There is no fast standard ways to synchronise the big amounts of data with the access protocols normally used for the HEP experiments.
- There is no well-established, secure, simple and popular way to access the data inside big institutions.

Copying ZEUS data from DESY DPHEP storage/ DESY dcache to RZG.

- Huge amount of data, order of hundreds of terabytes;
- Many-many-many files: 1.2 Mio;
- Very non-homogeneous set of files: different types and sizes, different handling in DESY.

Copying ZEUS data from DESY DPHEP storage/ DESY dcache to RZG.

- RZG supports srm, gsidcap, gridftp, webdav; visible and mountable worldwide.
- DESY DPHEP storage doesn't support any protocols; visible and NFS4 mountable on 2(!) machines nafhh-zeus in DESY.
- DESY dCache supports dcap; visible only from DESY, mountable only from ZEUS workgroup servers.

Copying ZEUS data from DESY DPHEP storage/ DESY dcache to RZG.

Therefore, the list of options to copy data from DESY to RZG:

- nafhh-zeus machines: DESY DPHEP $\rightarrow$ (NFS4 mount@nafhh-zeus) $\rightarrow$ (srm, gsidcap, gridftp, webdav, mount) RZG
- DESY computer: DESY dCache $\rightarrow$ (gsidcap, mount) RZG
- ZEUS workgroup server: DESY dCache $\rightarrow$ (mount@ZEUS workgroup server)$\rightarrow$ (srm, gsidcap, gridftp, webdav, mount) RZG

The srm is too slow (up to 1 minute per average file) and too CPU consuming, webdav-able utilities have authentication problems for writing: not able to handle VO extensions. Mount in most cases can be done via fuse and has the same problems. Those are not suitable at all.

Copying ZEUS data from DESY DPHEP storage/ DESY dcache to RZG.

The remaining options are (gsi)dcap and gridftp, however:

- (gsi)dcap-enabled utilities (dccp) are not able to create directories and do recursive copying, but have "do not overwrite" option, third-party copying depends on setup (impossible in this case).
- grid-enabled ftp clients (uberftp) are able to create directories and do recursive copy but doesn't have "do not overwrite" option, third-party copying depends on setup (impossible in this case).

No single tool provides enough functionality for the copying: in is necessary to use combination of dccp, uberftp and shell scripts.

Copying ZEUS data from DESY DPHEP storage/ DESY dcache to RZG.

The physical capabilities are limited:

- ZEUS workgroup server shut down.
- Publicly available DESY computers ("bastions" and "pal") are not intended for intensive I/O, attempts of large data transfers are banned.
- DESY Grid site uses virtual machines, the intensive I/O is not efficient and is recommended to be avoided.

Those are not suitable for data transfer. **The remaining options are only 2 nafhh-zeus machines.**

What we can learn from bit preservation problems

- Despite availability of multiple protocols and tools for data access in HEP, no single tool provides enough functionality for a simple copying. Possible solution would be to introduce the lacking functionality to dccp and uberftp. gsiftp-enabled rsync would be a tool of a dream.
- It is easy to get the CPU power and environment, but not capabilities for intensive I/O to work with old data. One of possible solutions would be a wider usage of third part transfers or dedicate some grid nodes for I/O only.
- In this moment providing access to the data with multiple protocols from any point in the Internet (like RZG is doing) helps to handle the data in the most efficient way. Technical restrictions should be imposed very carefully, with a proper reason, as many of them make even a simple data access unreasonably difficult;
- Trivia: less files is better. Even simple writing of log files to the ROOT file would reduce the number of files by 30%.

Copying H1 data from DESY DPHEP storage/ DESY dcache to RZG.

Exactly same problems as for ZEUS.

Conclusions

- The data preservation project at MPP is moving forward.
- Any help and commitment is welcome.

- DP for the active and future experiments will require more standartisation in the approach, tools, software and documentation.