ALEPH analysis preservation

Simone Coscetti - simone.coscetti@cern.ch
INFN Pisa
Engineering Ph.D. School ~ University of Pisa

Aims

* Fully functional stand-alone ALEPH installation.

* Use cases we wanted to cover:

    * a machine ready for cloud usage;

    * instantiate interactive machines on demand.

ALEPH environment

* Current situation using VirtualBox:

    * Installed a SLC4 distribution.

    * Needs for the ALEPH environment:

        * SLC4;

        * CERNLIB;

        * ALEPH software;

        * ALEPH data access.

ALEPH environment

* Last native environment of the experiment:

  * Linux RedHat 6.2;

  * direct access to CERN tape;

  * sw installed on AFS, heavy dependences on CERNLIB.

SLC4 and CERNLIB

* Why SLC4? It has been used few years ago for published analysis (last known official Aleph analysis use SLC4).

  * Also SLC5 is suitable for this purpose, but for preservation goals it is important to have a certified platform for analysis.

* CERNLIB rpm is available in the slc4 repository and no problem observed in the coexistence with SLC4 and ALEPH sw.

  * The integration of CERNLIB with SLC5 and ALEPH sw is more elaborated but not impossible.

How the VM was prepared

* SLC4 32 bit;

* CERNLIB available on yum repository;

* Made a local copy of the needed files stored under /afs/ cern.ch/aleph area.

* The system at the moment is totally independent from AFS:

    * moreover, it is independent from networking (can be used on a laptop while flying on a plane...) except DB access, needed when preparing list of file on which to run.

ALEPH data

* At the moment ALEPH data are stored on Castor at CERN.

    * Difficult to use it from a VM which can be anywhere in the world;

    * we are searching for a better solution that guarantees readiness and ease of use (e.g. external copies: ALEPH data occupy nowadays a little amount of space).

ALEPH data

* A relevant solution could be the implementation of a WebDAV server:

    * 10/20 TB in size to work with data as if they were in local.

    * SLC4 has a WebDAV native support.

* Other solutions are represented by latest technology like:

    * xrootd;

    * via castor with gLite command (lcg-cp, …);

   * sw wrapping with Parrot (CDF way, ALEPH works best with Posix).

Tested sw components

* The whole analysis chain has been reproduced:

  * kingal: event generator library;

  * galeph: MC simulation program for the detector;

  * julia: reconstruction program;

  * alpha: physics analysis package.

* compilation of new analysis and software works - not limited to replaying ancient analyses.
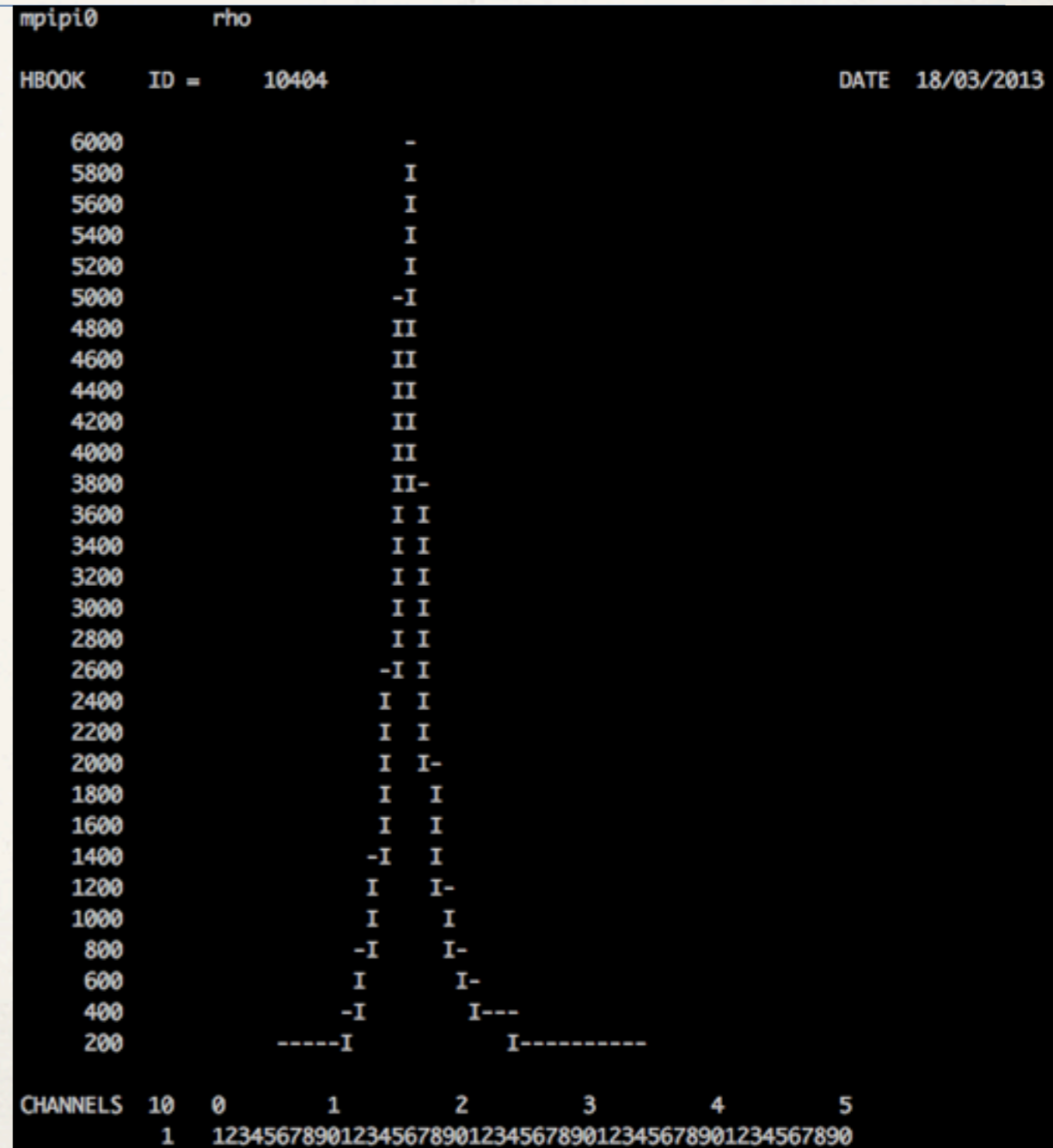
Tested sw components

* Interactive use - the system is suitable for interactive analysis:

* All the previous functionalities, plus the event display and PAW works
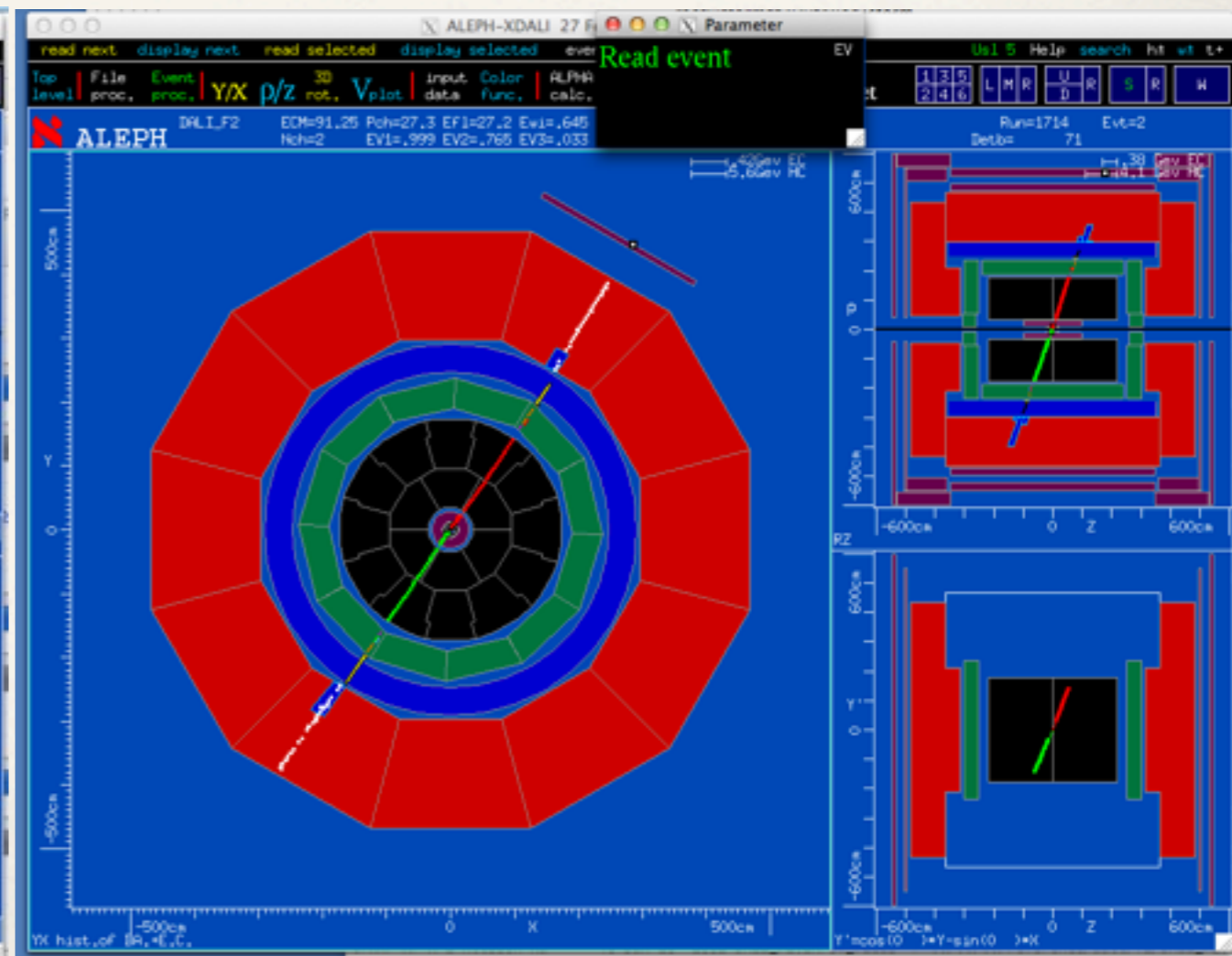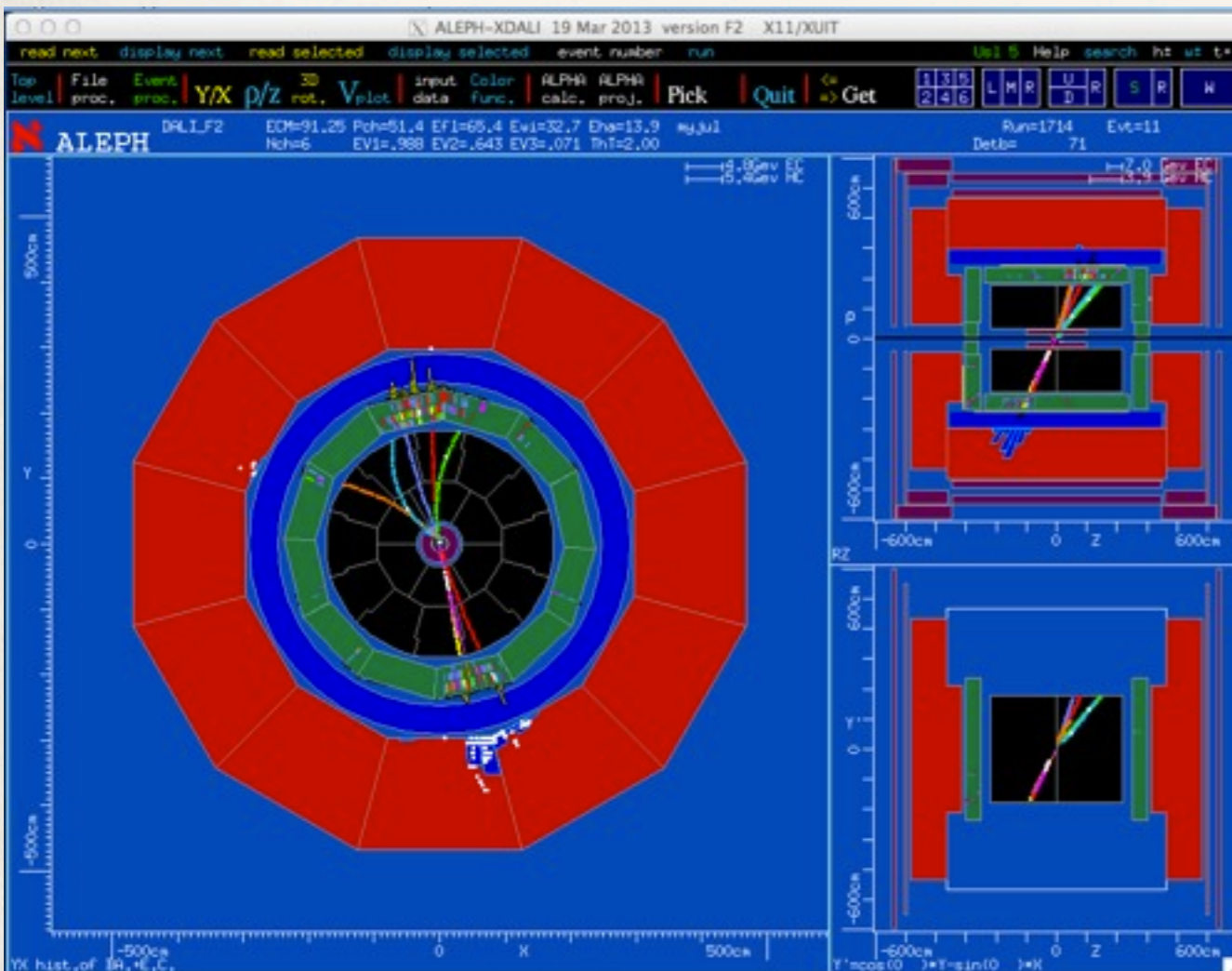
Applications

* The main goal could be the availability of these VMs in a cloud for jobs submission.

* At the moment a copy of the VM is available via srm at Pisa SE (~4 GB):

    * srm://stormfe1.pi.infn.it:8444/srm/managerv2?SFN=/cms/store/user/coscetti/aleph/

        * sl4_alephVM-disk1.vmdk and sl4_alephVM.ovf

    * user: aleph / password: aleph

Outstanding problems

* Event catalog not accessible since it resides on Oracle at CERN;

* Access to data (CASTOR);

* Many solutions are possible (under discussion: direct mount? WebDAV? …);

* The VM is as is:

  * we have not integrated cloud services yet;

  * works for local submissions.

Conclusions and To-Do

* Installed a SLC4 with fully ALEPH environment available:

    * no dependency from AFS and network issues;

    * no data db access.

* The whole ALEPH analysis chain has been reproduced:

    * the machine is ready and available for interactive usage.

* Need a coordinated approach with similar projects to estabilish (good) data access and scheduling of VM.