Preserving access to ALEPH Computing Environment via Virtual Machines

Simone Coscetti[1,2], Silvia Arezzini[2], Tommaso Boccali[2], Marcello Maggi[3]

[1]Università di Pisa, [2]INFN, sezione di Pisa, [3]INFN, sezione di Bari

Abstract

The ALEPH Collaboration took data at the LEP (CERN) electron-positron collider in the period 1989-2000, producing more than 300 scientific papers. While most of the Collaboration activities stopped in the last years, the data collected still has physics potential with new theoretical models emerging, and needing a check with data at the Z and WW production energies. An attempt to revive and preserve the ALEPH Computing Environment is presented; the aim is not only the preservation of the data files (usually called "bit preservation"), but of the full environment a physicist would need to perform brand new analyses. Technically, a Virtual Machine approach has been chosen, using the VirtualBox platform. Concerning simulated events, the full chain from event generators to physics plots is possible, and reprocessing of data events is also functioning. Interactive tools like the DALI event display can be used on both data and simulated events. The Virtual Machine approach seems suited for both interactive usage and for massive computing using Cloud like approaches. Studies are now moving from technical functionality tests (which are positively concluded), to tests and development on how to guarantee an easy and transparent access to ALEPH data in the virtualized platform.

Introduction

The data sets collected during many years of detector operations at particle accelerators offer unique opportunities for future scientific studies: new theoretical input, new experimental results and analysis technique, the quest for high-sensitivity combined analyses, the necessity of cross checks. In many cases high-energy physics (HEP) data sets are unique and they cannot superseded by data from newer generations of experiments. The cost of conserving this heritage through a long-term data preservation program would be small, compared to the costs of past experimental projects or to the efforts to re-do the experiments.

The HEP community will benefit from preserved data samples through re-analysis, combination, education and outreach. In most cases the basic (raw) data disappear in a period of five to ten years after the end of the data taking. The main reasons are rapid changes in storage technologies and computing and software systems.

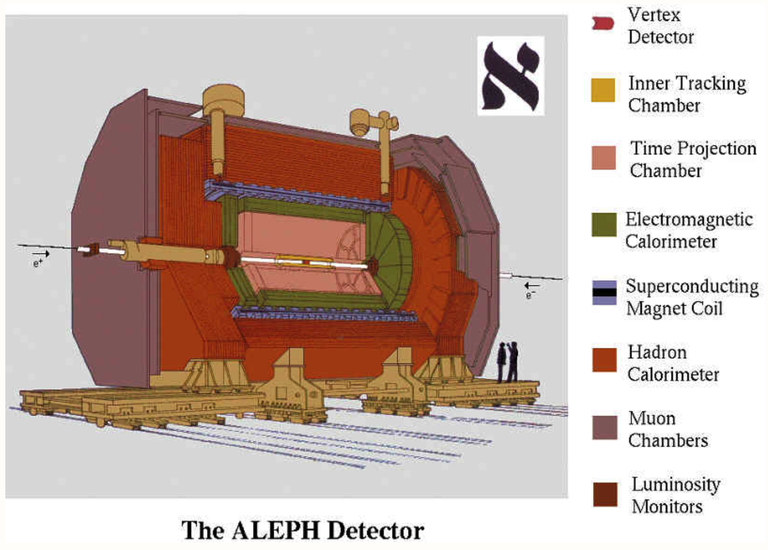
Long term preservation of HEP data is crucial to preserve the ability of addressing a wide range of scientific challenges and questions at times long after the completion of experiments that collected the data. In many cases, these data are and will continue to be unique in their energy range, process dynamics and experimental techinques. New, improved and refined scientific questions may require a reanalysis of such data sets. Some scientific opportunities for data preservation are: long-term completion and extension of scientific programs, cross-collaboration analysis, data re-use, education, training and outreach.

The ALEPH experiment

ALEPH was a particle physics experiment installed at the Large Electron-Positron (LEP) collider, at the CERN laboratory in Geneva/Switzerland. It was designed to explore the physics predicted by the Standard Model and to search for physics beyond it.

ALEPH first measured events in LEP in July 1989. LEP operated at around 91 GeV – the predicted optimum energy for the formation of the Z particle. From 1995 to 2000 (end of the work) the accelerator operated at energies up to 209 GeV, above the threshold for producing pairs of W particles.

This is a preliminary work on software and data preservation of the ALEPH experiment. The final target is represented by a fully functional stand-alone ALEPH installation. This ALEPH installation is able to work on the cloud, with a system to instantiate interactive machine on demand.



The ALEPH Detector

Machine Setup

A Scientific Linux 4 distribution has been installed on VirtualBox. A fully functioning ALEPH environment needs the CERNLIB, the ALEPH software and a full access to ALEPH data. The last native environment of the ALEPH experiment was based on Linux Red Hat 6.2, a direct access to the CERN tape was possible and the software was entirely installed on AFS with heavy dependences on the CERNLIB.

About the reasons of an SL4 distribution - SL4 has been used few years ago for the last official ALEPH analysis.

About the CERNLIB - The CERNLIB rpm is available in the SLC4 repository and no problems have been observed in the coexistence with SLC4 and ALEPH software.

About data and the access to them - CERN currently hosts archival ALEPH data on the Castor storage system. This entails a slow and complex access to them. We have searched for a solution that guarantees readiness and ease of use: data have been moved to a current generation disk system, and served via the WebDav protocol. SL4 supports *davfs2*, a Linux tool for connecting to WebDav shares as though they were local disks.

Cloud computing: a set of virtual machines like this is available on a cloud supplied by the INFN-Bari datacenter (thanks to ReCaS and PRISMA projects) for jobs submission. Such a system is implemented in an OpenStack instance. OpenStack is an open software consisting of a cloud operating system that can control large pools of compute, storage and networking resources throughout a datacenter.

Software Components

Physics Analysis is done by comparing data with simulated events. New physics would require the simulation and reconstruction to be able to produce new simulated data in the same format of the preserved data, and existing simulated events. For this reason all the components of the entire analysis chain together with the full documentation has been preserved:

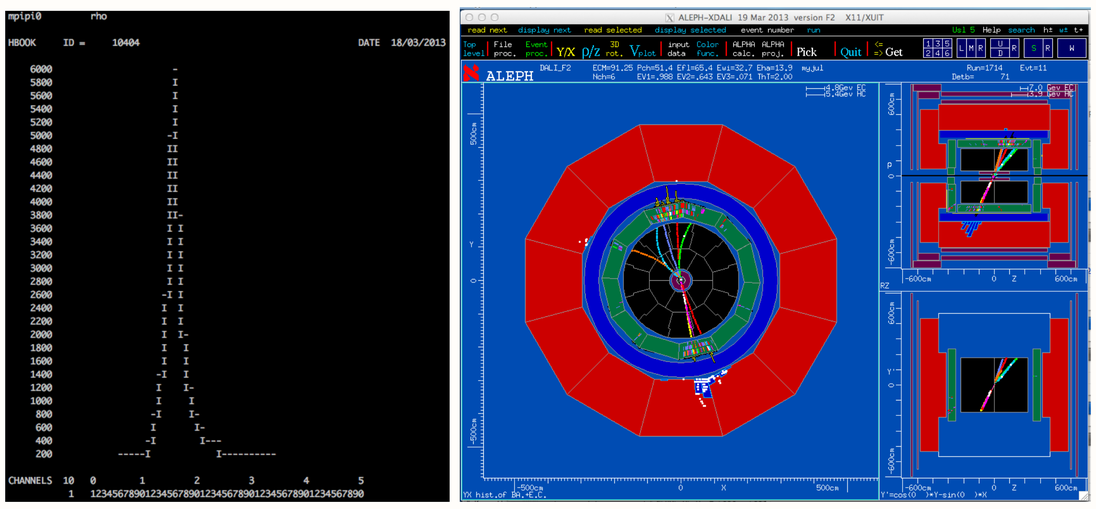
KINGAL the event generator library

GALEPH the MC simulation program for the detector

JULIA the recontruction program

ALPHA the physics analysis package

The compilation of new analysis and software works, not limited to ancient analysis. The system is suitable for interactive analysis, including all the previous functionality plus DALI (the event display) and PAW.



Output examples of KINGAL and DALI.

Applications

- The solution can be use to reproduce published analysis, and completely new studies.
- The solution includes a complete development environment, where software components at any step can be modified, recompiled, debugged.
- The whole software stack is available, for data and MC sets, for interactive and batch processing.

The work has been partially funded under contract 20108T4XTM of 'Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale (Italy)'

CHEP2013: October 14-18, 2013, Amsterdam, The Netherlands