NA2 Status: Computing and Analysis

Frank Gaede[1], Peter Wienemann[2]

January 09, 2007

Abstract

The objectives of the two NA2 tasks COMP and ANALYS are to provide a distributed grid-enabled computing infrastructure and the development of a common software framework for data analysis, simulation and data exchange. This memo describes the progress of these two NA2 tasks during the first twelve months.

1 DESY, Hamburg, Germany
2 University of Bonn, Germany

1 Introduction

A powerful computing infrastructure is an essential prerequisite to make progress in R&D activities towards a detector for the ILC. Therefore Network Activity 2 (NA2) comprises among its six tasks COMP and ANALYS.

The goal of COMP is to set up, during the first three years, dedicated computing clusters distributed over the contributing institutes. These clusters are interconnected and accessible from all EUDET partners by use of grid middleware. Institutions contributing to this task are University of Bonn/University of Freiburg, DESY and University of Tel Aviv.

ANALYS comprises the development of a common software framework to exchange testbeam data, to analyse and compare measurements, to simulate testbeam set-ups and the creation of a repository for experimental and simulation data. Commitments to this task come from University of Bonn/University of Freiburg, DESY and IPASCR (Institute of Physics, Academy of Sciences of the Czech Republic), Prague.

2 Computing resources

All contributing institutes have set up and are running grid-enabled computing clusters supporting the virtual organisations (VO) `ilc` and/or `calice` which are the organisational grid user units for ILC related work.

Freiburg University is presently running a grid computer cluster with approximately 100 CPUs and 10 TB storage which is partly shared with other VOs. The involved group who recently moved to Bonn University is starting to set up a cluster in Bonn. 30 000 € of EUDET money have been spent on computing nodes for this.

In 2006 DESY provided in total 55 890 CPU days of computing resources (status: September 2006). During that period 21 372 CPU days were used, 1889 of which by the VO `ilc` (9 %). In addition 9.6 TB of storage space are in use for ILC related work at DESY. To extend the existing infrastructure DESY purchased six EUDET dedicated SunFire X4100 computing nodes with four CPUs each for 23 010 € in total as a contribution to EUDET in 2006. The remaining 6 990 € went into two SunFire X5400 file servers (partly dedicated to ILC related work) with 24 TB disk storage each. The price of these file servers was 28 000 € each.

Tel Aviv University bought five dual core machines with 3 TB disk storage for a total cost of approximately 10 000 € from EUDET funds. These computers are included in their grid cluster and are configured in such a way that ILC related work is granted highest priority. In addition they started to integrate 28 PCs which can be used outside teaching hours. More resources will probably follow.

3 Structure of software framework

In order to avoid double work there is no EUDET/testbeam specific software being written but instead the EUDET work builds on top of the already existing ILC/LDC software. It is revised and its functionality is extended where needed for testbeam specific applications or other purposes within the EUDET range of tasks.

The basis of all software tools is the LCIO event data model and persistency framework [1]. It is already fairly complete and flexible. Nevertheless its data structures have been extended by a few classes (e. g. `TrackerRawData`, `TrackerData`, `TrackerPulse`) that are specifically suited for raw data as needed in the prototype studies on request by the user community.

The backbone for analysing and reconstructing LCIO data is the modular Marlin (**M**odular **A**nalysis and **R**econstruction for the **Lin**ear Collider) application framework [2] written in the C++ programming language. Its modules, called processors, sequentially process the event data by reading data from an input collection, process it and fill the results in an output collection which in turn can be processed by a subsequent processor. Which processors have to be called in which order using which parameters can be easily controlled by the user in XML steering files. The modularity and well defined interfaces between the processors ensures that different developers can work on different processors in parallel without interference. It also allows to plug'n'play with processors to try out e. g. different algorithms.

Funded by EUDET money a programmer has been hired at DESY. As a first task a number of helpful extensions have been developed for the Marlin framework. Examples are the addition of a sophisticated consistency check of steering files which is particularly valuable if long and complex steering files are used and the implementation of a graphical user interface to Marlin (MarlinGUI) which allows easy interactive modifications of steering files.

Further developments which were in particular driven by prototype and testbeam requirements are the Linear Collider Conditions Data (LCCD) toolkit [3] and GEAR (**Ge**ometry **A**PI for **R**econstruction) [4]. LCCD allows writing and reading of conditions data describing the detector status as function of time. It allows to tag data sets for later easy reference and to request data valid at a particular point in time. GEAR allows to access geometry information needed for the reconstruction of events. Thereby it is made sure that consistent geometry data are used throughout the reconstruction. The geometry information itself is stored in XML files. GEAR experienced significant extensions during 2006, mainly triggered by TPC and calorimeter prototype studies (e. g. a structure for rectangular TPC readout pad layouts). Another important development in this field is MokkaGear which allows to pass the geometry information used in the Geant4 [6] based full detector simulation to the reconstruction code via GEAR XML files.

4 Usage of provided hardware and software

Both the provided hardware and software are increasingly accepted by the ILC user community and are more and more displacing older group specific solutions.

The data processing scheme of the CALICE [7] testbeam software is based on the LCIO event data model. The raw data, as read from the experiment, is converted promptly into the LCIO format, which is then used for all further processing and analysis. The reconstruction algorithms are implemented as Marlin processors and the access to and storage of conditions data is provided by LCCD. The CALICE collaboration also moved to store all their testbeam data on grid storage elements making it easy for all collaborators to access and analyse the data on the provided grid computing resources. The grid is also used for Geant4 based simulations of testbeam set-ups that are needed to understand the recorded measurements.

In June 2006 an initiative was started by the LCTPC collaboration [8] to develop a completely new TPC analysis and reconstruction software (MarlinTPC) [9] to be used to analyse and reconstruct data which will be recorded by the large TPC prototype being built right now and to perform comparisons of data taken with many different small prototypes. This software is already designed with the given structure of LCIO, Marlin, LCCD and GEAR in mind. It is still under development but a usable release is expected in the course of the year 2007.

Another application of the grid resources provided by EUDET is the production of data sets with varied detector geometries which is essential to optimize the detector design. In total almost half a million events were processed on the grid and stored on grid storage elements such that they can be used by the ILC user community. In order to make it easy to find the available data sets, their properties and locations are stored in a Monte Carlo production database which can be easily queried via a web interface [10].

5 Conclusion

The tasks of COMP and ANALYS to set up interconnected grid computing clusters and to provide a common software framework for simulation and analysis of testbeam data is progressing well. The milestone to have the first stage of the computing network installed after ten months is fulfilled. Prospects are also good to fulfil the next milestone, namely to release a first version of the analysis framework after 18 months. The hardware and software offered by the COMP and ANALYS contributors are well accepted by the ILC community leading to a fruitful interplay between providers and users.

Acknowledgement

This work is supported by the Commission of the European Communities under the 6th Framework Programme "Structuring the European Research Area", contract number RII3-026126.

References

[1] LCIO home page, http://lcio.desy.de.
[2] Marlin home page, http://ilcsoft.desy.de/marlin.
[3] LCCD home page, http://ilcsoft.desy.de/lccd.
[4] GEAR home page, http://ilcsoft.desy.de/gear.
[5] Mokka home page, http://polywww.in2p3.fr/activites/physique/geant4/www/mokka/mokka.html.
[6] Geant4 home page, http://cern.ch/geant4.
[7] CALICE home page, http://polywww.in2p3.fr/activites/physique/flc/calice.html.
[8] LCTPC home page, http://www.lctpc.org.
[9] MarlinTPC home page, http://ilcsoft.desy.de/ilcsoft/ilcsoftware/marlintpc.
[10] MC production database, http://www-flc.desy.de/simulation/database.