COMPUTING AT H1 – EXPERIENCE AND FUTURE

G. ECKERLIN, R. GERHARDS, C. KLEINWORT, U. KRÜNER-MARQUIS

Deutsches Elektronen Synchrotron, Notkestrasse 85, D-22603 Hamburg, Germany

S. EGLI

Physikalisches Institut der Universität Zürich,
Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

F. NIEBERGALL

II. Institut für Experimentalphysik, Universität Hamburg,
Luruper Chaussee 149, D-22761 Hamburg, Germany

The H1 experiment has now been successfully operating at the electron proton collider HERA at DESY for three years. During this time the computing environment has gradually shifted from a mainframe oriented environment to the distributed server/client Unix world. This transition is now almost complete. Computing needs are largely determined by the present amount of 1.5 TB of reconstructed data per year (1994), corresponding to 1.2×10^7 accepted events. All data are centrally available at DESY. In addition to data analysis, which is done in all collaborating institutes, most of the centrally organized Monte Carlo production is performed outside of DESY. New software tools to cope with offline computing needs include CENTIPEDE, a tool for the use of distributed batch and interactive resources for Monte Carlo production, and H1UNIX, a software package for automatic updates of H1 software on all UNIX platforms.

1 Introduction

At the electron proton collider HERA at the DESY laboratory in Hamburg, positrons of 27.6 GeV momentum are collided with 820 GeV protons. A schematic view of the H1 detector[1] – one of the four big experiments operated at HERA – is shown in figure 1. The bunch crossing rate at HERA is 10.41 MHz giving a bunch separation of 96 ns. The H1 experiment is taking data at a logging rate of about 10 Hz or about 500 KB/s on average, resulting in a yearly raw data volume of about 5 TB. Data is reconstructed quasi-online, yielding another 5 TB of reconstructed data after physics oriented event classification. Raw and reconstructed data are kept on tape only. Part of the reconstructed event information is kept on disk in form of DST data. After software compression, the DST data amounts to about 300 GB[2,3]. These data volumes are a new challenge for experiments at HERA, even if these values are still orders of magnitudes lower than expected at future experiments, in particular at LHC.

2 Local Computing Environment

The H1 computing environment at DESY is shown in figure 2. The main computing

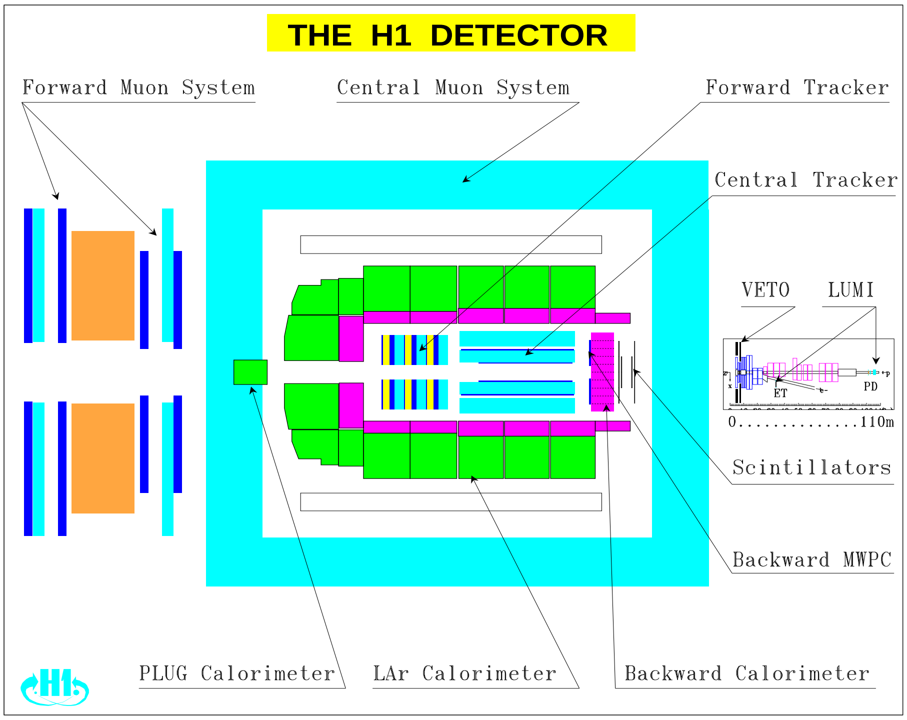

Figure 1: Schematic view of the H1 detector

facilities are provided by two SGI Challenge XL computers with 18 MIPS R4400 processors each, used for *online* (data logging, online reconstruction, re-processing, *dice1*) and *analysis* purposes (batch and interactive, *dice2*), respectively.

A *batch environment* has been set up on dice2, using NQS as the batch system. Job submission to dice2 is possible from many platforms in and outside of DESY. About 3000 jobs are submitted per week with at most 16 jobs running simultaneously. The same Challenge machine is also used for most of the *interactive computing* by the majority of the H1 members. Typical applications are N-tuple analysis, program development, submission of batch jobs, accessing information services, etc. About 100 to 150 users are logged in simultaneously and 300 users logon to dice2 at least once a day. It turned out that a smooth operation can not be provided by a single multiprocessor machine anymore. Therefore the H1 collaboration decided to join the workgroup server project initiated by the DESY computer center groups earlier this year[4]. In the autumn of 1995, the computing environment was therefore supplemented by 10 SGI Indy workstations, serving 5 to 10 users each. H1 aims to separated services for all aspects of computing, e.g. batch analysis, interactive analysis, program binary servers, home directory servers, etc. The Andrew File System (AFS) is used as the important merging part in the system.

Data that need frequent and simultaneous access by many users is stored on about 500 GB of disk space. The aim is to keep the data of at least two consecutive

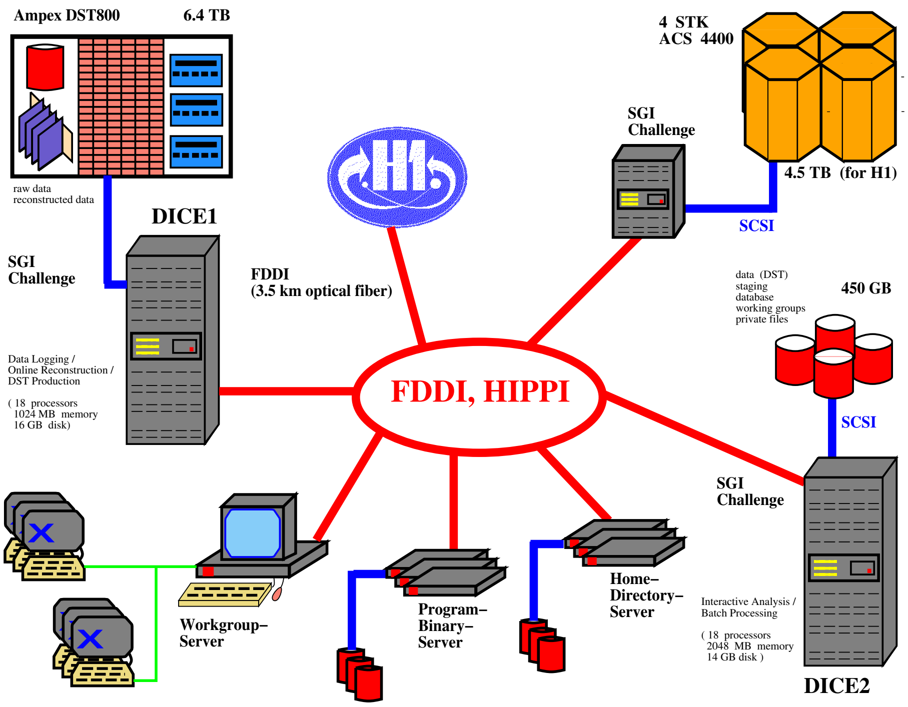

Figure 2: Computing environment of the H1 experiment

years on disk for combined analyses. Tape service is provided by an Ampex DST800 TeraStore mass storage system with 3 DST600 recorders, based on helical scan D2 technology, and 4 StorageTek ACS 4400 tape libraries, with about 15 % of its storage capacity and 25 % of its I/O capacity reserved for the H1 experiment. The capacity per Ampex tape amounts to 25 GB, the total robot capacity is 6.2 TB. Direct access to the StorageTek silos is provided by the Open Storage Manager (OSM) software[5]. The major components are interconnected via fast FDDI and HIPPI networks with a Netstar Gigarouter as the central router.

3 External Computing

An important part of the H1 computing concept relies on computing done outside of DESY. A substantial fraction of the physics analysis and most of the centrally organized Monte Carlo event production is performed externally on workstation clusters. All data is collected centrally at DESY in the StorageTek silo and the Ampex mass storage system on tape and on disk in the form of DST data. External sites keep private copies of (subsets of) data.

The tape media used for export and import are 3480/3490 cartridges (these are being phased out), Exabytes (which are recommended for small and medium size data), and DLT (which are recommended for medium and large data samples).

Four levels of Monte Carlo data are distinguished. Events at Monte Carlo generator level have a size of a few kilobytes. The event generation step is usually performed at DESY. It is the most time consuming GEANT tracking step (about $30\,s$ on a Challenge processor) which is performed at external sites, yielding events of about 100 KB size which are sent back to DESY. The digitization, trigger simulation, and reconstruction steps are executed at DESY (about $3\,s$ on a Challenge processor) and add another 100 to 150 KB to the event size. DST data are also generated for Monte Carlo events and contribute about 15 KB per event after software compression. The yearly data volume, assuming effectively 25 weeks of production, amounts to 2.5×10^6 (12.5×10^6) events or about 600 GB (3.0 TB) presently (planned for end of 1995). The expected data volumes for export are 500 GB per month (mainly DST and subsamples of data and Monte Carlo files) and for import 200 GB per month (Monte Carlo files).

The resources for re-processing at DESY allow at most one re-processing of the previous year's data. Therefore, an external re-processing of (raw) data is being investigated presently. This implies also a strong requirement on the data export and import facilities. All raw data will have to be exported from DESY and the reconstructed data will have to imported.

4 Distributed Computing

Another basic concept of H1 is distributed computing. To make full use of the CPU time available both at DESY and outside of DESY, the CENTIPEDE project was initiated. It provides a Monte Carlo production system with fully automized event distribution in a heterogeneous environment. CENTIPEDE is closely following the FUNNEL project[6] developed by the ZEUS collaboration. Unlike FUNNEL, CENTIPEDE is based on the PVM software package and is interfaced to the standard I/O package used in H1 (FPACK[7]).

Another important issue in a big collaboration such as H1, is code management. The H1 collaboration decided some years ago on CMZ since this is the only product that supports all platforms used within H1. All H1 software is maintained centrally at DESY at dice2 on CMZ files. Installation scripts are included in the respective CMZ file. A code distribution system (H1UNIX, H1VMS) is available for all UNIX platforms and VAX/VMS. It provides a set of shell scripts for a convenient update of the software. All machine dependent compiler and loader flags are isolated in one file. The update of CMZ files and libraries is done automatically, but it was considered essential that the final decision on production versions is performed by hand. With the introduction of AFS, an automatic central update of H1 software for all major platforms used within the H1 collaboration is planned. The various copies of libraries and executables will be kept on a program binary server at DESY.

5 Databases

All detector geometry and calibration data is stored in a purpose built database (MDB), which is accessed mostly by programs and offers limited interactive query capabilities. The database is based on a BOS direct access file on UNIX and VMS systems and on F-package[3,7] keyed access files on the IBM mainframes. The use of F-package ordered access files is in preparation. The current size is about 400 MB with a growth of 100 to 200 MB per year. Other database applications like run and Monte Carlo bookkeeping, slow control data, and personal data (e-mail, phone numbers, ...), which are dominated by interactive access queries, are being implemented into an Oracle database.

6 Conclusion

The three years since its commisioning have shown that the H1 experiment has successfully completed the transition from the central IBM mainframe to the UNIX world of distributed computing in the challenging computing environment at HERA. The consistent distribution of tasks to dedicated servers is well advanced but not complete yet. In particular, the separation of batch and interactive work turned out to be a key requirement for the offline computing. An upgrade of both the central batch capacity by another SGI (Power) Challenge and the number of H1 workgroup servers on the DESY site is envisaged for the first half of 1996.

With the development in progress on software tools and hardware installations, it is generally believed that H1 is well on the way for computing in the next millenium.

Acknowledgments

The authors gratefully acknowledge the tremendous efforts by the DESY computer center groups in providing a convenient and usable computing facility for H1. The authors also express their gratitude to their colleagues in the H1 collaboration for many fruitful discussions. The authors particularly wish to thank Vicky White and Alberto Santoro and all their co-workers from Lafex and Fermilab for organizing an exciting conference at such a splendid place as Rio de Janeiro.

References

1. I. Abt *et al.,* "The H1 Detector at HERA," DESY preprint 93-103 (1993), to be published in Nucl. Instr. Meth.
2. P. Fuhrmann *et al.*, "Data Logging and Online Reconstruction in H1", Proc. Int. Conf. on Comp. in High Energy Physics, San Francisco, USA, 1994, Ed. S.C.Loken
3. R. Gerhards *et al.*, "Data Storage and Data Access at H1", these proceedings
4. K. Künne, "Interactive and Batch Analysis at DESY", these proceedings
5. M. Gasthuber, "Distributed Mass Storage and Management Systems at DESY", Proc. Int. Conf. on Comp. in High Energy Physics, San Francisco, USA, 1994, Ed. S.C.Loken
6. B. Burow, "FUNNEL: Towards Comfortable Event Processing", these proceedings
7. V. Blobel, "The F-package for Input/Output", Proc. Int. Conf. Comp. in High Energy Physics, Annecy, France (1992), Eds. C.Verkerk and W.Wojcik