EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CERN–IT/99–7
November 26, 1999

Research and Development in Computing at CERN

professor Martti Tienari
Department of Computer Science & Helsinki Institute of Physics
University of Helsinki, FIN-00014 Helsinki
tienari@cs.helsinki.fi

Abstract

CERN is evaluated as working environment for researchers in computing sciences. For this purpose the main computing activities and research and development work in computing are surveyed from a computer science point of view.

1 Introduction

The author of this report worked for two months, October-November 1999, as a Corresponding Associate at CERN[1)] with the aim to explore from a Finnish point of view whether CERN would offer a suitable environment for the work of Ph.D students and post docs in computing sciences. For this purpose research oriented computing at CERN was surveyed, especially in the IT Division, but also multidisciplinary (e.g. physics/computer science) activities were taken in account.

Although CERN is a research centre of particle physics, there is a growing awareness that, as a by-product of building complex and advanced experimental research apparatus and measuring systems, valuable new insights and inventions are conceived in many other areas of science and technology. This has led to the establishment of a new Technology Transfer Division and also motivates research oriented work in other relevant areas, e.g., in computing. Of the c. 1600 papers published annually at CERN at least 100 are written about computing problems.

Work in CERN has led to many innovations in information technology. The best-known example is the invention of World Wide Web at CERN. The inventor of WWW, Tim Berners-Lee, did probably not receive as much support at CERN as his innovation would have deserved, but with the consent of his superiors he was able to develop the basic technology of WWW and demonstrate its usefulness. It satisfied a need in the international physics research community.

There are many ways a computing specialist could come to CERN and participate in its projects on information technology. There are many different types of positions available: scientific associate, CERN fellow (a post doc or Ph.D student), doctoral student (with shared supervision between CERN and the home university), technical student (e.g., to work on a MSc thesis project), summer student, or ordinary employee at CERN.

Some countries (at least Austria, Norway, Portugal, Spain, Sweden) are actively cultivating contacts at CERN for their non-physics students. Austria has a well-organised special programme to finance the work of doctoral students in applied sciences. Sweden has special arrangements for technical students.

Many researchers come to CERN to work as unpaid associates supported by their home institute, e.g., many Finnish physicists working at CERN are financed from Finland. For many areas of physics research CERN forms such an attractive environment that this kind of arrangement is well-motivated. The same could be also true for computer scientists and engineers.

2 Computing at CERN

The **IT (Information Technology) Division** in CERN is developing and running the general data processing and network environment. Its R & D is mostly fairly short-range, with a problem time-window of 3 – 12 months. This is a drawback from the point of view of a Ph.D student, who would normally prefer to participate in a more long-range, 2 – 3 year project.

The organisation of the IT Division is currently subdivided in groups as follows:

[1)] CERN is the European Laboratory for Particle Physics, founded in 1954, a joint venture of 20 member states. It is located on the border between France and Switzerland, just outside Geneva. Information of CERN and its activities can be found at http://www.cern.ch/.

Divisional Management and Secretariat
Application Software and Databases
Computing for Engineering
Central Infrastructure and Operations
Controls Group
Desktop Infrastructure Services
Information Process Technology
Internet Applications
Networking
Physics Data Processing
User Support

The manpower strength of the IT Division is 190 persons plus 75 outside visitors (fellows, associates, students etc). Research and development oriented groups in the IT Division are interested in getting competent visitors.

In **EP (Experimental Physics) Division** computer programs and databases are designed to handle the data generated in future experiments. There is obviously much cooperation with the IT Division, which helps EP Division teams in their work.

Currently several LHC (Large Hadron Collider) experiments, which should be operational in the year 2005, are under preparation. The problem time-window at EP Division is wider than at the IT Division, around 1 – 3 years, probably making the EP Division a more suitable working environment for Ph.D students than the IT Division. The whole atmosphere there is more problem-solving oriented than in the more service-oriented IT Division.

In **AS (Administrative Support) Division** much work is done in computing, too. There are 3 sizable groups, 50 computing-oriented people altogether: (1) Data Base Applications, (2) Systems and User Interfaces, (3) Scientific Information Services[2)]. The database applications are based on Oracle, much programming is done in Java. The applications are modern, e.g., electronic procurement and routing of documents in electronic form. They also have R & D activities in computing applications and employ students and fellows.

In **EST (Engineering Support and Technologies) Division** there is a group called Information Systems Support of 25 computing-oriented people supporting all engineering activities at CERN. The main working areas of this group are computer-aided engineering and engineering database support. A big integrated engineering information system CEDAR, in which Finland participates with a major web-interface project TUOVI, is a successful development result of this group.

CERN organises annually a two-week event, **CERN School of Computing**, to inform researchers in computing and physics about current developments in computing relevant to CERN. In 1999 the school was arranged in Poland, in 1998 in Portugal. More information can be found at http://www.cern.ch/CSC/.

The main titles of the 1999 school were: (1) Advanced topics (Quantum computing, Mass Storage Challenges for LCH, Current Mass Storage Systems at CERN),(2) LHC Experiments Data Processing and Data Communication Systems, (3) Software Building, (4) Internet Software Technologies (Distributed Computing Using Agents, Transaction

[2)] Will be moved to the new Technology Transfer Division.

Technologies, Advanced Web Software Topics). In 1998 the main titles of the school were: (1) Agents and Distributed Computing Technology, (2) Intelligent Monitoring and Control, (3) Petabyte Storage (4) Software evolution.

3 Research and Development in Computing at CERN

In the following, computing problems relevant for CERN are listed, classified by the main areas of computer science. This survey of topical problems has been gathered by interviewing various people in the IT, EP, AS and EST divisions.

Hardware Architecture

Super-computing at CERN is done mostly by computer farms, e.g., 1000 PC's processing in parallel event measurement data, each PC processing one event at a time. There are still open research problems in computing farms: how to organise and manage these farms etc

Computational Science

LHC experiments have complicated and heavy on-line and off-line data processing needs. They will use the huge PC farms mentioned above. There are various individual problems suitable for computer science researchers. A doctoral student of computer science is currently studying fault tolerance issues in the detector data acquisition system. A computer science oriented researcher is currently being searched for on-line measurement timing problems in LHC.

Database and Information Retrieval

In the new LHC (Large Hadron Collider) experiments the recorded data volume will be so large that the storage and retrieval of these data will be a major problem. These problems are suitable for computer scientists as research subjects. Several computer science doctoral students are already studying these issues.

The concepts and methods gained in the research of object oriented databases as well as distributed databases can be used in solving these problems and in optimising storage and retrieval of LHC data. There is also a physics part in this data organisation problem: one should be able to foresee, how the stored measurement data will be used in the future by the research physicists.

The CERN IT Division is cooperating with a Californian software company Objectivity in developing its object-oriented database software to become suitable for huge CERN databases of experimental results. Similar cooperation is taking place with Oracle's relational data base software.

EST division is also a good working environment for students interested in user interface design, distributed databases, data base searching and indexing.

Software Engineering

The experimental work at CERN requires very large software systems which are mainly constructed by physicists. A new trend in CERN computing is to use object oriented programming languages, mainly C++ but also Java, instead of the traditional Fortran. Software engineering methods used to manage large software systems are of increasing importance at CERN. A currently prevalent approach in designing software support for LHC experiments is to take the overall software architecture of the experiment

as the framework to manage the huge programming effort.

In software development one may find several schools of thought at CERN. The dominant school, which is supported by the IT Division, tries to use commercially available software tools as much as possible. There is also a minority group at CERN which believes in tailoring software to the physics needs and traditions, working from the bottom up. This group has created its own software framework based on C++ and object-orientation, which is called ROOT. For more information of this approach see http://root.cern.ch/.

There is some interest in compilers at CERN, especially in efficient optimised compilers for processor architectures with much internal parallelism. Portable GNU compilers are not as efficient as CERN people would hope in modern computer architectures, like Alpha or Merced.

Operating Systems and Networking

Usage of Linux on Intel PC is increasing at CERN. It has many advantages compared to the competing Microsoft technologies (Windows NT and Windows 2000). As an internal development effort there are projects to transfer Linux kernel and GLIBC-libraries to a new Intel/HP "Merced" processor (IA64). Linux device drivers are also locally developed for the new gigabit-networks which CERN is currently experimenting with.

The physics experiments at CERN require huge data transfer speeds and volumes. Thus an important area of development is to experiment with high-speed gigabit networks like GSN (Gigabyte System Network) or HIPPI-6400 and ultra-fast transfer protocols like ST (Scheduled Transfer). New approaches to control distributed storage, called SAN (Storage Area Network), can be implemented with these novel high-speed networks.

There is a group on Internet Applications at the IT Division. The current R & D projects of this group are indicative of what kind of research is currently done at CERN in general networking: traffic-analysis, monitoring end-to-end service quality, differential services on the Internet, collaborative multimedia and group-ware, email security, intrusion detection, user authentication and certificate authority. Although the Web Consortium W3C activities have been transfered mostly from CERN to INRIA (a big research institute in France) the extensive usage of web-technology at CERN still generates R & D problems to be solved locally at CERN. One topical subject, where a researcher would be needed, is "Distributed Authoring in the Web".

In the Desktop Infrastructure Services group there are ongoing long-range development projects in CERN workstation infrastructure management (Windows 2000 and Linux 2000 projects) as well as keen interest in computer and network security research.

4 Discussion

There are many computing challenges and problems at CERN, the analysis of which would benefit from the work of computer scientists. The CERN atmosphere is research-oriented and dynamic, one meets state-of-the-art and cutting-edge problems here. The international contacts, also with North America, are close and frequent. This is a place where, in addition to physicists, computer scientists would also feel themselves stimulated and well-connected with people interested in the same kind of problems.

Computer scientists coming to CERN should have an "applied" background, with skills and attitudes enabling them to become a productive member of some CERN team,

being able to solve relevant computing problems. A difficulty for an outsider is to make contact with people working on problems in his/her sphere of interest and locate problems at CERN which could be allocated, e.g., to a doctoral student. An orientation visit at CERN is in many cases necessary. The budget of CERN is gradually diminishing; in most cases special funding arrangements are needed to enable the work of an outside expert at CERN.

Senior computer scientists in member countries should have contacts to CERN and thus be able to inform young computer scientists of the possibilities offered by CERN. This is the way many technical students and doctoral students at CERN have found their way to Geneva. CERN could also be more active in distributing information about open computing research problems; this service can currently only be found in web pages describing Austrian special doctoral program.

CERN is very much dependent on collaborating with national research institutes in its mission. One way to promote computer science research at CERN, which would fit nicely with this tradition, would be to set up a research program in collaboration with CERN, which would devote its energies to some computing problem relevant for CERN and making it possible for Finnish computer scientists to contribute at a doctoral level. This kind of long-range project should be established by senior computer scientists in Finland in cooperation with CERN and the Helsinki Institute of Physics. This would follow the pattern set by the TUOVI project in industrial engineering software.

5 Conclusions

There are many working and educational possibilities for computer scientists at CERN. These possibilities should be exploited in Finland more than has been done so far. This multinational and multidisciplinary research centre is a good alternative for people in computing wanting to have an international working experience.

CERN is at its best for technical students visiting CERN for 6–12 months, who write their MSc thesis there. CERN can also be recommended as a good post doc place for a computer scientist wanting to stay abroad 1–2 years and to participate in applied computing-oriented R & D.

For doctoral students CERN can be recommended only if the stay is carefully planned in advance together with a professor at a Finnish university. A research subject in applied computing relevant for CERN, should first be located, for which competent supervision would be available at CERN and at the home university.

Working at CERN is currently not considered as a job opportunity by most computing students and researchers, because it is simply not known outside the physics community. The aim of this article is to spread information about CERN to the computing research community. The author is ready to give further information and help for Finnish computing scientists wanting to establish contact with CERN.