ФИЗИКА ЭЛЕМЕНТАРНЫХ ЧАСТИЦ И АТОМНОГО ЯДРА
2005. Т. 36. ВЫП. 1

УДК 004.75:539.125.523.348

COMPUTING NEEDS OF LHC EXPERIMENTS AND PROSPECTS FOR WORLDWIDE GRID FOR PARTICLE PHYSICS

P. Malecki, T. Szymocha, M. Turala
The Henryk Niewodniczanski Institute of Nuclear Physics,
Polish Academy of Sciences, Cracow, Poland

T. Bold
Faculty of Physics and Nuclear Techniques, UST AGH, Cracow, Poland

P. Nyczyk, A. Ozieblo
Academic Computer Centre, CYFRONET, UST AGH, Cracow, Poland

ФИЗИКА ЭЛЕМЕНТАРНЫХ ЧАСТИЦ И АТОМНОГО ЯДРА
2005. Т. 36. ВЫП. 1

УДК 004.75:539.125.523.348

COMPUTING NEEDS OF LHC EXPERIMENTS AND PROSPECTS FOR WORLDWIDE GRID FOR PARTICLE PHYSICS

P. Malecki, T. Szymocha, M. Turala
The Henryk Niewodniczanski Institute of Nuclear Physics,
Polish Academy of Sciences, Cracow, Poland
T. Bold
Faculty of Physics and Nuclear Techniques, UST AGH, Cracow, Poland
P. Nyczyk, A. Ozieblo
Academic Computer Centre, CYFRONET, UST AGH, Cracow, Poland

Several new experiments in particle physics are being prepared by large international community. They will generate data at the rate of 100–200 MByte/s over a number of years, which will result in many PBytes (10^{15} Bytes) of information. This data will have to be made accessible to a large international community of researchers, and as such it calls for a new approach to the problem of data analysis. Estimates of the computing needs of future experiments, as well as scenarios of overcoming potential difficulties, are presented, based on the studies conducted by LHC consortium and GRID computing projects. Short information on the operation of the LHC Computing GRID project is provided, together with the description of the first installations. Examples of large-scale Monte-Carlo simulations are also given.

Большим международным сообществом готовятся несколько экспериментов по физике частиц. В течение нескольких лет будут получены данные со скоростью 100–200 Мбайт/с, что составит большой (10^{15} байт) объем информации. Предполагается сделать эти данные доступными для большого международного сообщества исследователей и, как следствие, использовать их для нового рассмотрения проблемы анализа данных. Представлены оценки компьютерных ресурсов, необходимых для будущих экспериментов, а также пути преодоления возможных трудностей, основанные на разработках проектов консорциума LHC и GRID-вычислениях. Коротко, наряду с первичными инсталляциями, излагается информация о проведении GRID-вычислений на LHC. Приведены также примеры крупномасштабных моделирований методом Монте-Карло.

INTRODUCTION

The search for new, rare phenomena in particle physics, and/or precise measurements of selected processes calls for experiments at very high energies and very high rates (luminosities) [1]. A set of new experiments is under preparation for the Large Hadron Collider accelerator (LHC) at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland [2]. These experiments will further verify the Standard Model which governs the microworld, but they will also search for new particles — like the Higgs particle responsible for the phenomenon of mass — and for new states of matter — such as the quark–gluon plasma.

The four LHC experiments, ALICE, ATLAS, CMS and LHCb [3], are being developed at CERN by large, international consortium — up to 1500 persons from 150 institutions per experiment located on five continents. These experiments will produce an enormous amount of data, which will be partially filtered on-line, while still leaving a large portion for recording and off-line analysis. In addition, analysis of this data will require massive Monte-Carlo simulations (MC), and all generated events have to be stored as well. A rule has been established that all members of the consortium will have «equal access» to the data, which postulates the concept of «worldwide computing» — such demand presents a serious challenge for future experiments.

Several present experiments in the United States, Japan, and Europe, like CDF, D0, BaBar, Belle or COMPASS [4], are already producing data at a rate approaching 1 TByte per day, and their experience with data management and computing models will have considerable influence on future solutions.

1. DATA RATE AND VOLUME

To enable scientists to find the Higgs particle or detect some other rare phenomena, the collision rate at LHC will reach 10^9 events per second, each of which is expected to comprise numerous particles. In order to record all the necessary parameters (angles, momenta, energy, energy loss, Cherenkov and/or transition radiation) with the required granularity and precision, the LHC experiments will consist of millions of elements, and many signals have to be digitised with an accuracy of 10–16 Bit. The amount of particles and the required granularity and precision of measurements result in many bits of information per event. The collision rates and event sizes of past, present, and future experiments are shown in Fig. 1 [5].

The tremendous stream of initial data has to be somehow reduced to match the speed of recording media (disks), which is in the range of several hundred MByte per second, corresponding to 100–200 events per second — this means that a compression factor of about 10^7 has to be achieved. The problem of efficient data preselection is a separate issue which will not be discussed here; it should be noted, however, that it is one of the most serious challenges of LHC experiments.

Present experiments, which in many aspects serve as testbeds for developers of future system architectures and computing models, may be characterized by a set of parameters* shown in Table 1.

---

*Numbers in Table 1 are extracted from various papers and conference proceedings therefore are not always consistent; as such should be taken as raw estimates. The unit of computing power kSI2k corresponds to 1 Intel Xeon 2.8 GHz processor; kSI2k $\sim$ 100 SI95.

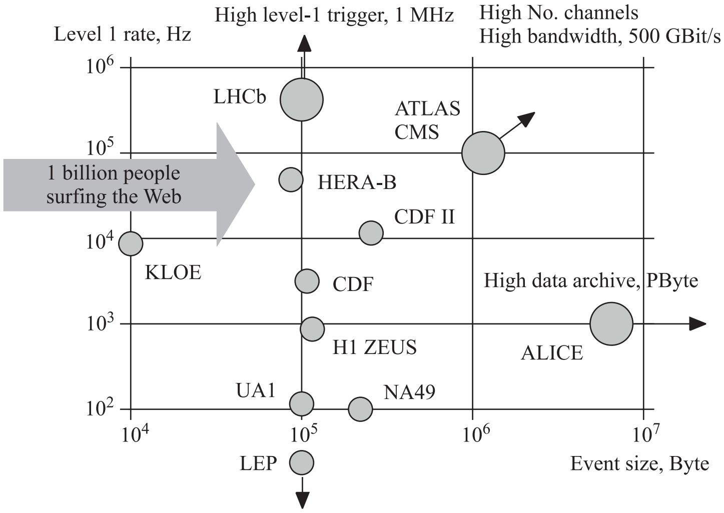

Fig. 1. Data rate and event size for past (UA1, LEP, CDF), present (KLOE, NA49, H1, ZEUS, CDF II) and future (ALICE, ATLAS, CMS, LHCb) particle physics experiments. Typical data flow in Web browsers is marked for comparison [5]

Table 1. **Characteristics of data flow and computing resources required by selected current experiments [6–9]**

	BaBar	Belle	CDF	COMPASS	D0
Event size, kByte	30	30–40	~ 250	30	~ 250
Events/y	$3 \cdot 10^8 – 2 \cdot 10^9$	$2 \cdot 10^9$	$\sim 10^9$	10^{10}	$\sim 10^9$
Raw data flux, MByte/s	~ 10	5–15	20	35	12–25
Storage, TByte/y	10–70	60	~ 100	300	~ 100
Total storage, TByte	~ 1000	~ 400	~ 1000	> 1000	~ 1000
Computing power, kSI2k	~ 30	~ 50	~ 10	~ 20	~ 10

2. COMPUTING ANALYSIS MODEL

The enormous volume of data, as well as the «equal access» requirement involving all collaborating institutions, force LHC collaborations to look carefully into possible solutions. The task is difficult, as the experiments are scheduled to run from the year 2007 onward. A set of groundbreaking technical proposals addressing the issue of LHC computing was prepared in 1996 [10]. These proposals specified the requirements and briefly addressed the questions of architecture, software and costs. It soon became clear, that in light of the rapid changes in the field of computer science, some intermediate studies were needed, to keep track of technical developments and to establish a common view on the matter of future computing for particle physics — one of the most essential problems was to agree on a common approach on both sides of the Atlantic. In 1998, a pilot research project called MONARC (Models Of Networked Analysis at Regional Centres) was launched by several European and American groups, the goal being to develop «baseline models», specify the main parameters characterizing the models' performance (throughputs, latencies) and verify the resource requirement baselines (computing, data handling, networks) [11]. As a result, a hierarchical model has been developed, with CERN being its Tier 0, large national laboratories forming Tier 1s, and smaller ones comprising Tier 2s — this is shown on Fig. 2.

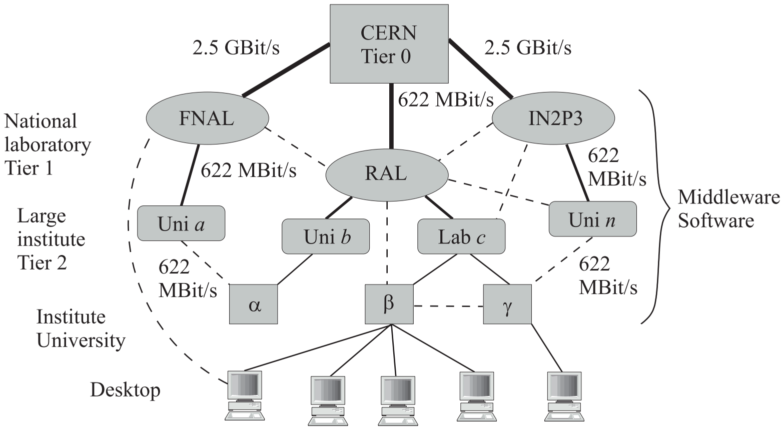

Fig. 2. The original architecture of distributed computing for LHC experiments, as defined by the MONARC project [11], followed a hierarchical structure; other connections (broken lines), were added later to schematically reflect the GRID concept

Thus, the original LHC computing infrastructure followed a hierarchical model. However, recent years have witnessed the emergence of a new concept of distributed computing, namely the Global Resource Information Database (GRID) [12]. The idea is to use distributed resources, processors, storage and network, as one system — this can be done through the use of very fast networks and adequate operating systems. It happens that the rate of developments in networking is twice the rate of similar advancements in hardware (processors and memories), which translates into one order of magnitude every five years [1]. Recent developments show that the network ceases to be a bottleneck for distributed systems, and even now 2.5–10 Gbit/s bandwidths are common for research, with much higher bandwidths expected in the near future.

When it comes to software, the task becomes more complex as it requires several nontrivial issues to be solved and standardized: authentication and authorization, security, data management and replication, brokerage of resources, system and job monitoring, friendly portal interfaces, accounting, etc. The Globus suite [13], developed in the USA, is widely accepted as a backbone for GRID computing, but it still requires adaptation for specific applications — several GRID projects are currently working on this [14–20].

At present, large HEP experiments implement a variety of computing models, trying to satisfy requirements principal for LHC collaborations, such as «equal access» for all collaborating institutions, user-friendliness of the system for physicists, and sufficient mass storage and computing power to store and analyze all the data taken by the detector.

The BELLE experiment has adopted a traditional HEP data and programming model [6] with raw data centrally stored and its reconstruction processing resulting in traditional DST (Data Summary Tape) production. Access to DST is available at any time, but relatively rare, with physics analyses done through mini-DSTs produced and distributed for particular tasks.

The BaBar experiment with 60 TByte/y data storing rate and with a goal of reaching 200 TByte/y bases primary event store on Objectivity, an OO database management system [7]. This choice of the data management model reflects considerably on the architecture of the OPR (On-line Prompt Reconstruction) farm of several hundred nodes as well as on several pseudoreal time reprocessing farms. The scale of BaBar allows one to expect that experience of that collaboration in many aspects of data storage and processing — such as those relating to problems of processing calibration prior to reconstruction of data samples, tagging and splitting event streams and efficiency of data distribution to remote sites will be of great importance for future LHC experiments.

Similarly, the COMPASS computing project [8] uses object-oriented database techniques for data storage and is developing a model reconstruction facility which will operate in parallel with the data gathering. A distributed computing farm of about 30 kSI2k, based on Fast Ethernet and Gigabit technologies, is the core component of the project.

Two large FNAL experiments, CDF and D0 [9], are highly distributed collaborations with scattered computing resources. Recently, both collaborations have intensified joint efforts on developing a data handling system, on large farms and large local networks and in particular on projects involving widely-distributed computing, including active participation in GRID projects. These activities are of particular value for the LHC as the running experiment is the most natural and demanding referee of new solutions.

3. GRID TECHNOLOGY AND PHYSICS

A central concept behind the GRID is structuring computing resources to be assigned to Virtual Organizations (VOs). Such VO groups resources are owned by «traditional» organizations (like institutes, universities, research centres) to solve a certain computing problem. Resources are expressed as computing power, disk/tape storage capacity, network throughput and specialized equipment, such as medical scanners, etc. The Globus Toolkit (GT) [12, 21] provides the means to construct VOs devoted to solving a variety of computing problems.

Currently, the most advanced developments in the area of HEP are based on GT, which is still in the development phase, however its version 2 is widely used by several GRID projects and considered stable. Current efforts are focused on a technologically advanced version 3, implementing the so-called Open GRID Services Architecture (OGSA) [22].

The most critical issues in GRID computing involve authentication and authorization. These problems are addressed by Globus in the following way: communication in the network of GRID enabled machines is encrypted by using the X.509 [23] system operating a Public Key Infrastructure (PKI). Each user is equipped with their own certificate which works as a «passport». This certificate, valid for some years, is electronically signed by a regional Certification Authority (CA), which vouches for its credibility. The user referenced by such a certificate may then generate temporary certificates, called proxy certificates, valid for some short period of time (12 h by default). Finally, the proxy certificate is used to encrypt communications and for user authentication. Authorization is solved by VO membership — which works as a «visa». With a valid certificate in hand, the user can submit jobs to Computing Elements (CE) being part of VO resources and transfer files from/to Storage Elements (SE) also available to that VO. Basic tools for file transfer and job submission are also provided by Globus.

In order to extend Globus functionality, more advanced projects are being developed, namely CrossGRID [15], European DataGRID (EDG) [14], intercontinental DataTAG [16], GRID Physics Network (GriPhyN) [17], Particle Physics Data GRID (PPDG) [18] or the International Virtual Data GRID Laboratory (iVDGL) [19], all fully or partially devoted to HEP applications. An overview of these projects is presented in Table 2. Arguably the most important of those, the European DataGRID (EDG), led by CERN, is largely devoted to feasibility studies of GRID technology for LHC computing. Some additional middleware and new services driven by LHC computing needs are developed by EDG. The European collaboration of research institutes within EDG has created the so-called «testbed», a geographically distributed collection of resources, where new developments are tested and real processing performed. Also testbeds primarily established for other projects, like CrossGRID or NorduGRID [20], are joining this infrastructure to provide an unprecedentedly large computational platform

Table 2. **Overview of GRID projects supporting HEP applications**

Project	Participants	Main objectives	Involved HEP experiments
EDG [14]	21 partners from Europe	Build the next generation computing infrastructure providing intensive computation and analysis of shared large-scale databases; HEP, Biomedical, Earth Observation	ALICE, ATLAS, CMS, LHCb
CrossGRID [15]	21 partners from 11 European countries	Development of software environment for interactive applications: Biomedical, Flood prediction, HEP, Pollution forecasting	ALICE, ATLAS, CMS, LHCb
PPDG [18]	11 research centres from US	Develop, acquire and deliver vitally needed GRID-enabled tools for data-intensive requirements of particle and nuclear physics	ATLAS, BaBar, CMS, DZero, STAR
GriPhyN [17]	18 US research centres	Develop GRID technologies for scientific and engineering projects that must collect and analyze distributed, PByte-scale datasets: HEP, Laser Interferometer Gravitational-wave Observatory (LIGO), Sloan Digital Sky Survey (SDSS)	CMS, ATLAS
iVDGL [19]	13 participants from US, Japan, Australia, Europe	Provide a unique laboratory that will test and validate GRID technologies at international and global scales, that will serve experiments in physics and astronomy. Sites in Europe and the U.S. will be linked by a multi-GBit/s transatlantic link funded by the EU DataTAG project	All HEP experiments are concerned
DataTAG [16]	22 partners from Europe and US	Transoceanic network studies, optimization of bulk data transfer, interoperability between Europe and US GRIDs testbeds	All HEP experiments are concerned

on which one could run several applications, including physics simulation and analysis.

Efficient interconnection of research centres in Europe is available thanks to the GEANT network [24] and, on a worldwide scale, by many other initiatives such as the Global Terabit Research Network (GTRN) [25]. Extensive review of the requirements and the status of national and international networks could be found at [26]. Recent tests performed by the EU DataTAG project [27] have demonstrated that the data transfer rate in a high GBit/s range is accessible: on Oct. 1, 2003, 5.44 GBit/s sustained rate, single TCP/IPv4 stream, was achieved between US (Caltech) and Europe (CERN), which meant the transfer of 1.1 TByte in 26 min (1 CD of 680 MByte/s).

A user wishing to execute jobs on the GRID has to prepare a job definition readable by GRID services. If raw Globus software is chosen for that purpose, the job must be described in a special Resource Specification Language (RSL). EDG and CrossGRID also allow jobs to be specified using the relatively more advanced Job Description Language (JDL). Job description contains information about which executable to run, the required input files, the desired machine architecture, memory, disk space, etc. A sample JDL file, with comments, is presented below:

```
Executable = "/bin/ls";  # executable located on host
Arguments = " -la";
Stdoutput = "ListOfFiles.txt";
StdError = "stderror";
# files which are downloaded by client when the job concludes
OutputSandbox = {"ListOfFiles.txt","stderror"};
RetryCount = 6;
# user requirements for batch system and operating system
Requirements =other.LRMSType == "PBS" && other.OpSys=="RH 6.2";
```

In this example one can see that even in heterogeneous resource collections users can enforce certain rules regarding the platform, operating system, etc. One may also notice that resources are not specified directly (by name) — this choice is not supposed to be of concern to the user. In fact, dedicated services, such as the EDG Resource Broker (RB), decide where to submit user jobs to optimise the overall performance of the system. Such abstraction of resources frees the job from dependence of local configurations and allows it to be executed in a highly dynamic environment. All functionality related to locating the desired files, closest replicas [29], etc., is delivered by appropriate GRID services. Enabling application for execution on the GRID, i.e., «gridyfying», seldom requires changes in the source code. The preferred solution is to «wrap» the application in a set of scripts providing GRID-compatible IO [30]. Naturally, modification of application source code to make it GRID-capable is also possible. The GT embodies an application programming interface (API) for all necessary GRID operations, albeit using it requires substantial expertise and tends to be rare in HEP applications. Code instrumentation is, on the other hand, common whenever parallel libraries (MPI) are in use. New standards are currently emerging in this area (i.e., MPICH-G2 [31]) and they are also being applied to HEP applications [32].

The idea and practical elaboration of the GRID concept could not have shown up at a more opportune moment from the point of view of experimental particle physics. Massive requirements for computing power and for maintaining PByte mass storage systems are the hallmark of LHC-era experiments and have led to formulation of computing models (MONARC) based on the worldwide distributed architectures, in remarkable convergence with the ideas behind the GRID. Physicists can again, as has often been the case in past decades, actively participate in the development of computer technologies. Experimental particle physics not only brings a strong demand for computing power but — equally important — necessitates the use of gigantic storage with efficient and sophisticated access, considerably enriching the original idea of the GRID.

4. LHC COMPUTING GRID

The progress in information technology, as well as experience with using commodity hardware for data analysis (a modern PC with Linux appears to be a good platform for the analysis and modelling of physics events), coupled with better understanding of the LHC experiments, necessitated a review of requirements and possible solutions. Such an in-depth review took place in the years 2000–2001 [33]. The most representative requirements concerning data volumes for one general-purpose LHC experiment are shown below (in reality there may be some discrepancies between experiments, due to hardware differences and different physics being involved).

Original rate of events	$\sim 10^9$ event/s (luminosity 10^{34} cm$^{-2}\cdot$s^{-1}, collision rate 40 MHz).
Events to disk (preselected)	$\sim$ 100 event/s (elaboration of the trigger — initially 270 event/s).
Event size	$\sim$ 1 MByte/event (better understanding of the apparatus — initially about 2 MByte).
Volume of data per day	$\sim$ 10 TByte/d (better understanding of the apparatus — initially about 50 TByte/d).
Volume of data per year	$\sim$ 1 PByte/y (better understanding of the apparatus, data for a 100-day run).
Monte-Carlo data	$\sim$ 1 PByte/y.

Requirements for all four LHC experiments concerning the Tier 0 and Tier 1 computing infrastructures (see Fig. 2) are presented in Table 3. Table 4 presents requirements concerning the necessary bandwidth [1, 26] which would allow access to the data and subsequent analysis and exchange of information.

The GRID concept has been recognized as a viable solution for the LHC computing infrastructure, as it offers better sharing of resources and loads, which

Table 3. **Computing resources required by LHC experiments**

	ALICE	ATLAS	CMS	LHCb
Tier 0 (CERN)				
CPU, kSI95	824	690	820	225
Disk pool, TByte	535	410	1143	330
Automated tape, TByte	3200	8959	1540	912
Shelf tape, TByte	—	—	2632	310
Tape I/O, MByte/s	1200	800	800	400
Cost 2005-7, MCHF	18.1	23.7	23.1	7.0
Tier1				
CPU, kSI95	234	209	417	140
Disk pool, TByte	273	360	943	150
Automated tape, TByte	400	1839	590	262
Shelf tape, TByte	—	—	683	55
Tape I/O, MByte/s	1200	800	800	400
Number of tier	4	6	5	5
Cost av, MCHF	7.1	8.5	13.6	4.0

Table 4. **Network bandwidth requirements (BW) for particle physics communities (in MByte/s)**

	1998	2000	>2005
Physicist	0.05–0.25	0.2–2	0.8–10
(peak BW)	(0.5–2)	(2–10)	(10–100)
University group	0.25–10	1.5–45	34–622
Regional centre/laboratory	1.5–45	34–155	622–5000
Transatlantic link	1.5–20	34–155	622–5000
Link to the laboratory where experiments take place	34–155	155–622	2500–5000

could reduce the underlying costs (the GRID concept is schematically marked in Fig. 2 by broken lines and «middleware» software).

The scale of the LHC computing infrastructure is so large that nothing similar has ever been attempted before. Therefore, it was deemed necessary that a pilot project involving a substantial fraction of the final system be built and tested. Fulfilling this requirement, the LHC Computing GRID (LCG) project was launched in 2002 [34], with the parameters of a CERN prototype, as shown in Table 5.

Table 5. **LCG requirements for GRID infrastructure**

	2002	2003	2004	2005
Processor farm				
No. of 2-CPU systems installed	400	530	660	800
Estimated total capacity, SI95	33 000	49 900	73 800	110 200
Disk storage				
No. of disks installed	400	530	990	1 200
Estimated total capacity, TByte	44	80	198	240
Tape drives				
Total capacity, achievable MByte/s	350	450	600	800
Automated media				
Total capacity, TByte	100	200	400	600

The concept of future computing infrastructures is gradually departing from the hierarchical MONARC model, and recently the idea of a «computing cloud» [35] has been introduced (see Fig. 3). The «cloud» would contain all the necessary resources, connected by fast networks and controlled by GRID middleware — such a solution could optimize the computing infrastructure with regard to access time and cost. The «cloud» could be used by all LHC experiments, and resources would be allocated based on the virtual organization (VO) scheme — of course, such a global approach would require adequate management.

In order to realize the international dimensions of the LCG project, its management was complemented by an LCG Deployment Board [36], which includes representatives of all LHC experiments and all major regional computing centres engaged in the analysis of HEP data, with the goal to discuss and make, or prepare, the decisions necessary for planning, deployment and operation of the LCG. The management of the LCG Deployment Board has created five working groups, whose aim was to define:

- security rules (certificate authentication, authorization, accounting);
- policies for identifying and sharing resources;
- a definition of the services to be provided by Tier 0, Tier 1, and Tier 2 centres;
- plans and schedules for deployment;
- a common software environment (LCFG, OS, GRID Services. . .);

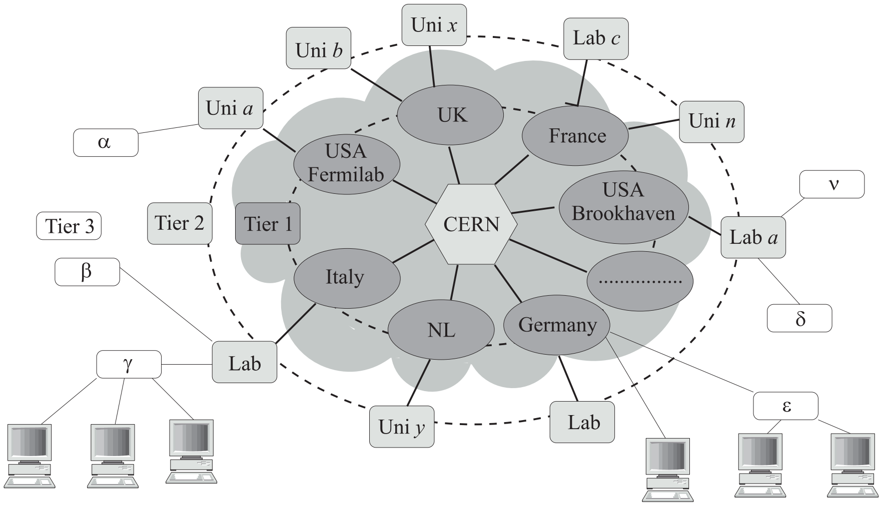

Fig. 3. The GRID concept has stimulated the development of a new architecture for LHC computing, namely a «computing cloud». A set of specific resources can be assigned to a particular experiment via a «virtual organization» scheme [35]

- operating and reporting standards and procedures;
- support for a common infrastructure.

Key documents (or drafts thereof) concerning the above topics were created in the first half of 2003 [36].

The basic operating model says that sites are responsible for providing computing resources that meet the requirements of the experiments — which, in turn, have the duty to collaborate with site managers to ensure efficient resource usage. The Regional Centres are responsible for managing their local operations support and for contributing to and interfacing with the global LCG-1 operations support. Middleware as well as any tools or packages required to be installed at LCG-1 sites must be fully tested and certified by the LCG Deployment team before local system administrators are asked to install them.

The LCG will be one of the first projects to deploy GRID technology on a very large production scale. This will require Virtual Organizations and institutes to define and agree on robust security policies and procedures that will enable the building and maintenance of «trust» between the various bodies involved. Users will register once, and only once, to be granted authority to use LCG-1 resources. An important feature of the registration approval process is that robust checks will need to be made to confirm the user's identity and that he/she is entitled to have access to the resources, VOs, groups and roles requested.

Fig. 4. Tier 2 LCG-1 cluster at ACC Cyfronet, Cracow, consisting of a set of dual Xeon 2.4 GHz processors, connected via 100 MBit/s Ethernet, with a 1 GBit/s uplink

Operating the large LCG computing infrastructure in a production mode will require substantial effort. Based on the need of a 24/7, 365-days-a-year support, the idea is to have three different «master» support centres, spread all over the world, in three different time zones. These support centres will provide a single point of contact to the clients and to the local GRID operations. The central support staff is the only interface to the GRID user in case of a problem, though it is not expected to solve that problem by itself. Likewise, the central User Support cannot provide training or integration support, but can coordinate both.

The decision to «allocate» an adequate number of resources and to support the particular experiment's data requirements lies with the LHC, the Computing RRB, and other CERN oversight mechanisms that balance physics goals against overall resource requests. It will be very difficult to implement consistent accounting across all LCG-1 sites during this first phase, but a well-defined accounting model should be available for the second release and tested during the LCG-1.

In parallel with the work on policies and operating procedures, LCG-1 software is being developed and tested: it is based on EDG middleware version 2.0, released by the DataGRID project [14], VDT package [38], GLUE scheme [39] and some LCG extensions and local modifications. The deployment of LCG-1 software and the formation of the LCG computing infrastructure commenced in spring 2003. In September 2003, the LCG-1 included 14 sites (mainly Tier 1s), from three continents: several European countries, Japan and Taiwan, Russia, United States. The actual status of the LCG infrastructure can be monitored at http://mapcentre.rl.ac.uk/. It is planned that the LCG infrastructure will reach stability and efficient operation at the beginning of 2004, so a massive production of physics simulation results could start at that time.

The Polish Tier 2, which right now includes only a small cluster of Intel processors running Linux at ACC Cyfronet, Cracow (see Fig. 4), has been connected to the system with the help of a Tier 1 operated by the Forschungszentrum Karlsruhe and tested by the LCG operation team. It is planned that polish Tier 2 will include infrastructure in Cracow and Warsaw each consisting of about 128 processors and 20 TByte storage serving mainly local physics communities.

Table 6 gives information on approximate resources available to LCG in 2004 in different countries.

Table 6. **Share per country of selected resources available to LCG-1 testbed at the end of 2003 [37]**

Country	CPU, kSI2k	Disk, TByte	Tapes, TByte
CERN	700	100–160	1000
Czech Rep.	30–60	5–6	2–5
France	150–420	24–81	160–540
Germany	207–305	40–44	62–74
Holland	30–124	1–3	12–20
Hungary	. . . –70	. . .	. . .
Italy	507–895	60–110	100
Japan	125–220	40–45	50–100
Poland	48–86	5–9	5–28
Russia	102–120	12.9–30	26–40
Taipei	40–220	5.8–30	30–120
UK	486–1780	55.3–455	100–300
USA	80–801	25–176	20–1741
Switzerland	18–26	4–5	20–40
Spain	150	30	100
Sweden	. . . –179	. . . –40	. . . –40

It is also expected that substantial help in the operations will be provided by the newly-approved EU project called EGEE (Enabling GRID for e-Science in Europe) [40], which considers LCG as its first pilot application.

5. LHC SIMULATIONS ON THE GRID

All four LHC experiments are vitally interested in GRID development. Their computing teams are actively taking part in GRID exercises and perform the so-called «stress tests» of the GRID infrastructure. This activity is mostly performed to test the system, but useful MC data samples are also being processed. Such trial runs are commonly known as Data Challenges (DCs).

It is foreseen that the ATLAS experiment will produce 1.3 PByte per year of raw data. Together with reconstructed events and MC data, 10 PByte per year will be required. This amount of data cannot be processed at CERN. The ATLAS experiment has initiated DCs to validate its computing and data model, integrate with LCG software and produce the required MC samples. In order to carry out the production phases in reasonable time, the distributed resources of participating institutes are involved. So far, Data Challenge 1 has been completed, with 50 million events (from which 11 million were complete physics events) generated and passed trough detailed detector simulation. 24 TByte of data were produced in 170 kSI2k-days and 8 TByte of MC samples in 2.5 kSI2k-days. The involvement of different partners in DC1 Phase 2 (simulation of LHC pile-up events) is shown in Fig. 5 as an example; more details can be found in [41]. Each site taking part in ATLAS production had to be validated. The procedure was based on processing test samples and comparison with reference results. About 100 histograms were created to test every aspect of the detector response simulation. The Cracow ATLAS group using Cyfronet cluster took part in DC1 Phase 1 and 2 and the site was validated successfully in 2002.

The main physics aim assigned for the Cyfronet cluster covered simulation of the ATLAS detector response for 50 k events Higgs boson decaying in channel HW→ $\mu\nu gg$ (at 400 GeV).

This ATLAS production was data-intensive. With the Cyfronet cluster, one event took 180 s to be processed on one CPU. Since each event takes up 2 MByte of space, 38 processors can produce 36 GByte/d. The output files were transferred to CERN tape storage (CASTOR). The bbftp tool provided multithreaded, quick and reliable transfer of hundreds of large files (exceeding 1 GByte per file).

The CMS collaboration has initiated the so-called «stress test», which will verify the portability of the CMS software environment and produce data for physics studies. The test was performed on the US Integration GRID Testbed where about 1.5 million events were generated in about 1 month and on the European DataGRID Testbed where more than 250 thousand events were produced

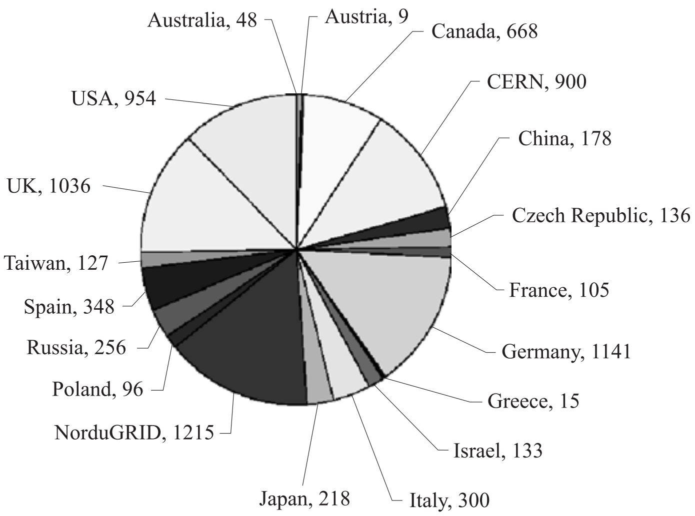

Fig. 5. Number of normalized processors per country participating in ATLAS DC1 Phase 2 (1 unit corresponds to 1 Pentium III 500 MHz equivalent to 190 SI2k) [41]

in some 3 weeks. Over 300 CPUs and 3 TByte of disk storage were allocated for these tests. The secondary goal of the stress test was the production of 1 million simulated events in less than 5 weeks.

In the second half of 2003, a Pre-Challenge Production will be performed in order to prepare over 50 million event samples required in Data Challenge 2004. In DC04, plans are to reconstruct digitized events at Tier-0 sites at a rate corresponding to 5% of the rate of LHC running with full luminosity (25 Hz, 50 MByte/s) using LCG-1 resources and services. More details can be found in [42].

The ALICE experiment has performed four Data Challenges since 1998. Each of them introduced new technology available at the time as well as more recent version of existing components. Since the end of 2001, the ALICE experiment has been using its own implementation of GRID services — ALIce ENvironment (ALIEN) [43]. The main objectives for DC IV, which ran between June and December 2002, were as follows:

- a scalability test for the Data Acquisition System (DAS);
- data transfer inside the DAS at 650 MByte/s minimum, sustained for a few hours;
- data recording to Permanent Data Storage at 200 MB/s minimum, sustained for seven consecutive days;
- 200 TByte of data being recorded to Permanent Data Storage.

These objectives were successfully met. For more details see [44].

The LHCb collaboration performs annual Data Challenges. The latest (DC2003) focused on the preparation of simulated event statistics for the trigger Technical Design Report and production of 30 million minimum bias events, 10 million generic *b*-decay events and 50 k–150 k events per 30 decay channels. There were 18 centres participating and 80% of CPUs were placed outside CERN (CERN — 20%, Great Britain — 12.8%, Italy — 11.5%, France — 9.0%). For the near future, DC2004 forsees:

- a robustness test of the LHCb software and production systems;
- a computing model test;
- distributed analysis;
- incorporation of the LCG application area software into the LHCb production environment.

The DC2004 will be run by the production manager at CERN in close collaboration with LHCb production site managers. In over 3 months over 220 million events will be produced (10 times DC2003). 50% of the CPU resources will be accessed via the LCG-1 prototype. In the more distant future, Data Challenge 2005 plans to conduct the following:

- mimicking LHC data acquisition;
- testing performance of the High Level Trigger, reconstruction and streaming software;
- evaluating the feasibility of a full reconstruction at 200 Hz in the on-line filter farm.

CONCLUSIONS

It appears that once-insurmountable data analysis problems related to LHC experiments, can, in fact, be solved due to continuous and rapid developments in information technologies, hardware and software. New ways of distributed computing are made available through progress in networking. It is difficult to make predictions regarding technology and market situation in the years 2006–2007 (and onward), which is when the experiments will start to run; however, past experience (LEP experiments saw progress far outstrip expectations), as well as newly-emerging global concepts of distributed computing (Grids) can provide a measure of optimism.

Acknowledgements. Materials used in this article came from many sources. In addition to the printed papers we have explored presentations of I. Bird, F. Gagliardi, R. Jones, M. Kasemann, M. Mazzucato, A. Putzer, L. Robertson, and W. von Rueden.

Deployment of LCG Tier 2 testbed in Cracow is due to interest and support of M. Bubak, J. Niwicki, and M. Noga of ACC Cyfronet UST-AGH and of H. Marten, R. Maschuw, K-P. Mickel, and P. Hasselmeyer of FZK Karlsruhe.

Editorial help of P. Nowakowski in the preparation of this article is admitted.

This work is supported in part by the grants of the European Commission, IST-2001-32243 and the Polish State Committee for Scientific Research 620/E-77/SPB/CERN/P-03/DZ 110/2003-2005, 620/E-77/SPB/5.PR UE/DZ 465/2002-2004, and 112/E-356/SPB/5PR UE/DZ 224/2002-2004.

REFERENCES

1. *Kasemann M.* Computing and Data Analysis for Future HEP Experiments // Plenary talk at XXXI Intern. Conf. on High Energy Physics, Amsterdam, July, 2002; http://www.ichep02.nl/index-new.html
2. The Large Hadron Collider Accelerator Project. Report CERN/AC/93-03(LHC). 1993;
   http://lhc-new-homepage.web.cern.ch/lhc-new-homepage/
3. Large Hadron Collider. The Experimental Programme. Report CERN. 1993;
   ALICE Technical Proposal. http://consult.cern.ch/alice/Documents/1995/01/abstract
   ATLAS Technical Proposal. Report CERN/LHCC/94-43. 1994;
   http://atlas.web.cern.ch/Atlas/TP/tp.html
   CMS Technical Proposal. Report CERN/LHC 94-38. 1994;
   LHCb Technical Proposal. http://lhcb-tp.web.cern.ch/lhcb-tp/
4. *BaBar Collab.* The BaBar Detector // Nucl. Instr. Meth. A. 2002. V. 479. P. 1;
   Abashian A. et al. (BELLE Collab.). The BELLE Detector // Ibid. P. 117–232;
   Abe F. et al. (CDF Collab.). The CDF Detector: An Overview // Nucl. Instr. Meth. A. 1988. V. 271. P. 387–403;
   COMPASS CERN/ SPSLC/96-14. SPSLC/P 297. 1966;
   D0 Collab. The D0 Detector // Nucl. Instr. Meth. A. 1994. V. 338. P. 185.
5. *Delfino M., von Rueden W.* // LHC Computing Workshop, INP, Cracow, Nov. 2000.
6. *Katayama N. et al.* Belle Computing Model. Elsevier Preprint;
   Adachi I. et al. Computing System for the Belle Experiment. physics/0306120.
7. *Gaponenko I. et al.* The BaBar Database: Challenges, Trends and Projections.
   http://www.ihep.ac.cn/∼chep01/paper/4-013.pdf
8. *Lamanna M.* The COMPASS Computing Farm Project // Proc. of the CHEP 2000 Conf., Padova, Italy, Feb. 2000. P. 576.
9. *Wolbers S.* Computing Experience from CDF and D0 // CHEP2003, March 24, 2003;
   White V. et al. The D0 Data Handling System. http://www.ihep.ac.cn/∼chep01/paper/4-050.pdf
10. ATLAS Computing Technical Proposal. Report CERN/LHCC/96-43. 1996;
    http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/TDR/html/Welcome.html
    CMS computing technical proposal. Report CERN/LHCC/96-45. 1996.
11. MONARC Project. http://monarc.web.cern.ch/MONARC/
12. *Foster I., Kesselman C.* The GRID: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999.
13. The Globus Project. http://www.globus.org/
14. European DataGRID Project. http://www.edg.org/
15. EU CrossGRID Project. http://www.eu-crossgrid.org/
16. DataTAG Project pages. http://datatag.web.cern.ch/datatag/

17. GRID Physics Network Project pages. http://www.griphyn.org/
18. Particle Physics Data GRID Project pages. http://www.ppdg.net/
19. International Virtual Data GRID Laboratory Project pages. http://www.ivdgl.org/
20. Nordic Countries GRID, NorduGRID Project pages. http://www.nordugrid.org/
21. *Foster I., Kesselman C., Tuecke S.* The Anatomy of the GRID — Enabling Scalable Virtual Organizations // Intern. J. High Performance Comp. Appl. 2001. V. 15(3). P. 200–222; http://www.globus.org/research/papers.html
22. *Foster I. et al.* The Physiology of the GRID: An Open GRID Services Architecture for Distributed Systems Integration. http://www.globus.org/research/papers/ogsa.pdf
23. RFC 2459 — Internet X.509 Public Key Infrastructure Certificate and CRL Profile. http://www.faqs.org/rfcs/rfc2459.html
24. EU Dante and Geant Project. http://www.dante.net/
25. Worldwide Next Generation Internet to be Established for Research and Education. http://www.gtrn.net/
26. *Newman H.* Networks for High Energy and Nuclear Physics and the ICFA SCIC // CHEP03 Conf. for Computing in High-Energy and Nuclear Physics, La Jolla, California, USA, March 24–28, 2003.
27. *Martin O.* DataTAG Project Update // Cracow GRID Workshop 2003, Cracow, Poland, Oct. 28, 2003.
28. EDG Tutorial. http://hep-proj-grid-tutorials.web.cern.ch/hep-proj-grid-tutorials/
29. *Cameron D. G. et al.* Evaluating Scheduling and Replica Optimization Strategies in OptorSim IRST Technical Report 0306-29 // Proc. of the 4th Intern. Workshop on GRID Computing (Grid2003), Phoenix, USA, Nov. 2003.
30. *Hoschek W. et al.* Data Management in an International Data GRID Project // The 2001 Globus Retreat, San Francisco, USA, Aug. 9–10, 2001.
31. *Karonis N., Toonen B., Foster I.* MPICH-G2: A GRID-Enabled Implementation of the Message Passing Interface // J. Parallel Distributed Comp. (JPDC). 2003. V. 63, No. 5. P. 551–563.
32. *Martinez-Rivero C. et al.* Interactive High-Energy Physics Application // Annual CrossGRID Project Workshop, Santiago de Compostella, 2003; http://grid.cesga.es/crossgrid/crossgrid.html
33. Report on the LHC Computing Review. Report CERN/LHCC/2001-004; http://lhc-computing-review-public.web.cern.ch/lhc-computing-review-public/
34. The LHC Computing GRID Project. http://lhcgrid.web.cern.ch/LHCgrid/
35. *Putzer A.* Worldwide Computing Model // ATLAS NCB Meeting, CERN, Sept. 19, 2001;
    Jones R. W. L. ATLAS and GRIDPP // GRIDPP Collab. Meeting, Edinburgh, Nov. 5, 2001;
    Jones R. W. L. ATLAS Computing and the GRID // Proc. of the VIII Intern. Workshop on Advanced Computing and Analysis Techniques in Physics Research, Moscow, June 24–28, 2002; Nucl. Instr. Meth. 2003. V. 502. P. 372–375.
36. LHC GRID Deployment Board. http://lcg.web.cern.ch/LCG/PEB/gdb/default.htm
37. *Mazzucato M.* Resources Available for LCG in Regional Centres // Presentation to closed C-RRB session, April 15, 2003; http://lcg.web.cern.ch/LCG/C-RRB/
    Mazzucato M. GDB Achievements and Resources in Regional Centres. http://lcg.web.cern.ch/LCG/C-RRB/oct-03/mirco-C-RRB-21-10-03.ppt
38. Virtual Data Toolkit. http://www.lsc-group.phys.uwm.edu/vdt/
39. GRID Laboratory Uniform Environment (GLUE). http://www.hicb.org

40. EGEE (Enabling GRID for e-Science in Europe) Project.
http://egee-ei.web.cern.ch/egee-ei/2003/index.html

41. ATLAS Data Challenge 1 Report. http://atlasinfo.cern.ch/Atlas/GROUPS/SOFTWARE/DC/DC1/DC1-Report-V3[1].0-22092003.pdf

42. *Grandi C.* Plans for the Integration of GRID Tools in the CMS Computing Environment // Computing for High-Energy and Nuclear Physics, La Jolla, California, USA, March 24–28, 2003;
http://www-conf.slac.stanford.edu/chep03/register/administrator/papers/papers/THCT004.PDF

43. Alice Environment Project pages. http://alien.cern.ch

44. *Anticic T. et al.* Challenging the Challenge: Handling Data in the GBit/s Range.
http://arxiv.org/abs/hep-ex/0305044