Istituto Nazionale di Fisica Nucleare LABORATORI NAZIONALI DI FRASCATI
SIDS-Pubblicazioni

LNF–10/05(IR)
February 19, 2010

FEASIBILITY STUDY FOR THE REALIZATION OF A SCIENTIFIC COMPUTING SERVICE AT THE LABORATORI NAZIONALI DI FRASCATI OF INFN

Report of the Scientific Computing Service Working Group

M. Benfatto, S. Bianco, F. Bossi, V. Chiarella, S. Dell'Agnello,
R. de Sangro (WG Chair), P. Di Nezza, M. L. Ferrer, E. Pace, L. Pellegrino,
R. Ricci, F. Ronchetti, F. Terranova, E. Vilucchi
INFN, Laboratori Nazionali di Frascati, P.O. Box 13, I-00044 Frascati, Italy

Abstract

This document contains the final report of the working group set up to study the feasibility of a Scientific Computing Service at the INFN's Laboratori Nazionali di Frascati.

The goal of the working group was to determine the location, infrastructure, hardware, software, manpower and funding profile needed to support the many computing activities of the experimental groups operating in the laboratory, including an ATLAS tier2 centre and an analysis farm for ALICE.

Contents

1 Executive Summary

This document contains the final report of the working group set up to study the feasibility of a Scientific Computing Service (SCS) at the INFN's Laboratori Nazionali di Frascati.

The mandate given to the working group was to determine the location, infrastructures, hardware, software and manpower needed to set up a Scientific Computing Service, together with a funding profile to complete it. The most important facts contained in the report and the conclusions reached by the working group are summarized here.

Presently the Laboratori Nazionali di Frascati run a centralized computing service (C&NS) which guarantees all the basic Intranet/Internet services such as network access and security, e-mail, web-mail, printing, authentication, web pages, centralized installations, DHCP and also provides a large support to several administrative services (i.e. Data-WEB, Servizio Informativo), used not only by the Laboratory but also the whole INFN organization.

The centralized resources that the C&NS dedicates to scientific computing are presently very limited. On the other hand, in the laboratory there exist several small clusters of computers which are independently operated by the groups that use them. This situation has historical origins we need not discuss here, and presents an unnecessary multiplication of human and infra-structural resources.

These clusters, situated in different areas of the laboratory, are managed by physicists or engineers (an estimated 2.8 FTE), expert in computing, whom could otherwise devote more of their time to their principal physics or technological research activity. The clusters' sizes, however, are small enough that, if they were part of a coherent system housed in one place, could be easily managed by one expert IT engineer instead.

Also the KLOE experiment at Frascati runs an independent computing centre which provides all the computing power, storage and data archiving needs of the collaboration. This system is housed part in the Computing Centre ground floor and part in the KLOE experiment building near DAFNE. The envisaged upgrade of the KLOE computing system can be accommodated easily in the present sites with some upgrades of the cooling and power systems, which can in part be shared with those of the SCS. The working group conclusion was that the KLOE computing need not, and should not, be included in the evaluation of the SCS project.

In the laboratory there are research groups that are either already heavily involved, or plan to increase their involvement in the next years, in the LHC computing. Among these are the ATLAS group, which is already running a proto-TIER2 presently housed in the Computing Centre Building, and the ALICE group which will be involved in the data analysis in the longer future. The Atlas group has already passed a first scrutiny from

Table 1: Evolution of computing resources required in the next 5-6 years.

	July 2009			July 2011			2012-2014		
	kSI2k	TB	KW	kSI2k	TB	KW	kSI2k	TB	KW
HEP	289.0	100.0	24.0	1640.0	828.0	75.6	2240.0	1148.0	109.1
Astro-Particle	133.0	20.0	23.0	205.0	35.0	4.2	205.0	35.0	4.8
Nuclear	72.0	1.2	13.5	911.0	261.0	31.4	1460.0	530.0	59.7
Theory	12.0	8.0	1.5	380.0	6.0	13.5	380.0	6.0	13.5
Total	495.0	129.2	62.0	3136.0	1130.0	129.2	4285.0	1719.0	186.9

INFN, and is pending the final approval of the Frascati site as a full fledge Atlas-TIER2, whereas Alice is planning to ask for a large increase in computing power on site in the near future.

Beside these larger (computing wise) enterprises, at Frascati there are also many smaller experimental groups which will greatly benefit from the availability of computing power deployed on site. There are BaBar, CMS, CDF-2, NA62, Super-B, Opera, Nautilus and the space based experiments Pamela and Lares.

The Frascati theory group is planning to expand their lattice QCD calculation activity, with a growing need to access large amount of parallelized computing power. The group has applied to INFN for funding an increase of the hardware presently dedicated to the group activities, and is looking forward to more substantial upgrades in the future.

The survey work by the working group has also exposed the growing computing needs of the Accelerator Division group, most notably for the design of new accelerators. The computing power required by beam dynamics calculations and machine parameters optimizations are significant, and in the past they have often been performed using computing resources obtained in kind outside the laboratory. Moreover, the software used to perform these calculations involve also a tedious and time consuming work of license procurement, managing and installation, which takes up lots of time and is presently often performed by the individual end user.

The conclusion of the working group is therefore that there would be a clear benefit from the institution at Frascati of a computer centre dedicated to scientific computing, large enough to be able to host in the same hall all the presently existing computing clusters plus the new hardware that could be acquired in the future, all managed by a relatively small group of dedicated IT engineers.

Table 1 illustrates the predicted 5-year evolution with time of the computing resources needed at Frascati, grouped by scientific line. The increase in the first two years for the CSN-1 reflects the likely hypothesis of the approval of the Atlas TIER2 sometimes next year, whereas the increase for the CSN-3 reflects a possible upgrade of the Alice and Panda systems starting in 2011.

The working group determined that a realistic estimate for the actual time needed to implement the first stage of the plan would be about 12 months. Assuming a prompt approval of the project, we chose July 2009 as the hypothetical start up date for the SCS, which is about 1 year from the release of this report.

All the hardware needed can be housed in the present Computing Centre building, once an appropriate reconfiguration of the internal layout and an upgrade of the power and cooling systems is completed.

The working group has found that there is the possibility and the convenience to stage the development of the SCS so as to follow the natural evolution of the actual requirements with time.

The first stage, which is mainly focused around the deployment of the Atlas TIER2, will be the most important as it will have to include all the infra-structural work on the Computing Centre hall needed to support the final hardware configuration. In this way, the additional work needed to adjust the system to all subsequent computing hardware installations will be done, as they occur, with minimum disruption of the SCS operations.

The working group has performed a feasibility study on the cooling and power supply upgrades necessary to the present Computing Centre hall in order to host the new hardware; the results are summarized in the additional documentation attached to the present report. Based on these studies, an outline of a time profile of the necessary work and cost of the deployment of the SCS, together with the manpower requirements to manage it, is summarized in Table 2. The Table includes the extra cost, dominated by the electric power bill, of operating the SCS with respect to the present situation. The first year extra cost of 18 k€ is due to the expected small increase of computing power to be acquired in the next 12 months.

Table 2: SCS funding (in k€) and manpower (in Full Time Equivalent) time profiles.

	1/7/2009	1/7/2010	1/7/2011	2012-2014
Cooling System	150.0	0.0	110.0	0.0
Electrical System	60.0	0.0	90.0	0.0
Computing Hardware (Networking)	45.0	0.0	20.0	0.0
Total Infrastructure Upgrades	255.0	0.0	220.0	0.0
Operating Extra Cost	18.0	72.0	125.0	211.0
Grand Total	273.0	72.0	345.0	211.0
Manpower (FTE)	2.8+1.0	2.8+1.0	2.0	3.0

In the startup phase it is required and expected that a level of manpower support equivalent to what has been given to run their own clusters will be granted by the various groups to the newly forming SCS. This means that the research groups will contribute about 2.8 FTE to transfer all the machines from their current location to the SCS hall and start up the service.

This kind of support will gradually switch from system management and hardware installation to the support of the research groups software, if needed, while these support tasks will be taken over by dedicated IT engineers. This additional manpower is estimated at 1 FTE for the first 1-2 years of operations, growing in time as the size of the SCS increases, as shown in Table 2. It is also strongly recommended that the C&NS continues its present level of support to scientific computing ($\approx$0.5 FTE) well into the development phase of the SCS, to insure a liaison with the existing computing infrastructure.

It is important to point out in closing, that a seizable fraction of the infra-structural upgrades of the Computing Centre hall would have to be done in any case as they are needed by the KLOE experiment when they upgrade their system for the foreseen high luminosity run, and by the C&NS to insure sufficient redundancy to critical services which the group presently provides to the central INFN organization.

2 Introduction

This report contains the result of a study carried out in spring 2008 by a group of LNF staff members to verify the possibility to create a structure within the Laboratory to support the computing needs of the various research groups active in the Frascati Laboratories.

2.1 Working Group Mandate and Objectives

The mandate given to the Working Group was to determine the location, infrastructures, hardware, software and manpower needed to set up a *Scientific Computing Service*, together with a funding profile to complete it. The facility should be dimensioned to meet the computing needs of as many as research groups in activity in the Frascati Laboratories as possible, from the largest LHC groups (Alice, Atlas, CMS, LHCb) to the theory group, from DAFNE to SPARC and SPARX and other Accelerator Division groups, over a time scale of up to 5 years from now.

The goal was therefore to understand what are the computing needs of the LNF research groups and their evolution with time, and determine what kind of infrastructure and manpower from the Laboratory would be required so support it.

Since INFN has well defined procedures for investing funds in computing hardware through the scientific and technical scrutiny of its Commissione Calcolo e Reti (national Committee for Computing and Networking) and of the other national Scientific Commissions (CSN1, CSN2, etc.), we deliberately did not perform any precise evaluation of the requests presented by the individual groups in terms of their scientific need or congruity.

We did evaluate, however, their plausibility using basic criteria of reasonability. For example we judged the request put forth by the Alice group a reasonable one, as the CPU and storage they ask for is of the same order of magnitude of other (already approved) Alice TIER2's, is at present foreseen as a possible evolution over the next 5 years, and comes from a group with a strong analysis team. We did not however, investigate or determine whether the amount of resources requested is actually needed and congruous on scientific and technical grounds.

We have therefore explicitly neglected to evaluate the cost of the computing hardware (CPU and storage) of projects requiring large amount of resources (i.e., at the TIER2level), as these would need (and have to seek) higher level INFN approval and thus are beyond the scope of our mandate.

We limited the cost estimates to those civil infrastructures (hall, cooling, electrical power, safety) and support computing infrastructures (network switches, etc.) which pertain to the Laboratory. Since the evaluation of the necessary infrastructures needed for the support of a computing centre is driven by the requirement of providing redundant electrical power and sufficient ambient cooling to an adequately dimensioned hall, we have evaluated the number of machines needed (CPU, disk servers, network switches), calculated the space required to house them (number of racks, etc.) and summed up their total power consumption (electrical and cooling).

2.2 Document Overview

We have divided the document in four chapters. The present introductory one in which we briefly outline the motivation and scope of the work and put it in the context of the LNF activity and the INFN organization. We also briefly recall the variegate multitude of research activities carried out at LNF.

In chapter 3 we present a survey of all the *existing* computing resources presently used at LNF by these groups; we collected information about the deployed hardware (CPU measured in kSI2kfor computing power[1], disk/tape storage measured in TB), software (OS, tools, etc.), infrastructure and manpower invested to support it.

In chapter 4 we surveyed the *additional* computing resources that the existing research activities require now and may do in the future. We conducted a poll among the various research groups to asses what kind of growth in computing needs we should anticipate on a time scale of $3 - 5$ years.

The outcome of the poll comprises well defined computing projects (such as the Atlas TIER2, which has already gone through INFN's CCR referral, has been approved as *Proto-TIER2* and is likely to be funded to a full size TIER2in the near future), as well as less defined requests which for now represent mere desiderata which may or may not find INFN's approval and funding in the future. However, our approach has been to put together all the requests, including less firm ones, with the objective to determine what could be a (reasonable) maximum total of computing resources the future Scientific Computing Service should be able to accommodate and manage, if they all turned out to be acquired over time. The idea is that the infrastructures should be scalable over time, without any major overhaul, as the requirements for computing grow up to the level foreseeable today. In chapter 5 the proposal of a viable realization of the Scientific Computing Service is presented, based on the data collected and analyzed in the preceding two chapters. It contains a description of the hardware and software resources necessary to fulfill the computing needs of the local research groups, where they could be housed, what infrastructures and manpower are needed, what kind of support is expected from

[1] We used this units as most of the older CPU are only rated in these units. We have therefore used them throughout this document

the existing Computing & Networking Service, and finally a time-line for its completion, with a funding profile.

2.3 Computing Activities within the INFN

Computing plays a fundamental role in any research activity, particularly so in the ones carried out by INFN. INFN has long since been active in the development of computing to achieve its own mission, through the actions and projects carried on within, and financed through, the Commissione Calcolo e Reti (CCR, INFN national commission for computing and networking). In doing so INFN has also played over the years a leading role in the development of Italy's research network infrastructure (GARR). The need for increasingly large computing power in terms of CPU and storage by the experiments of the LHC era, their increasing complexity, and therefore the increasingly larger fraction of the cost of the experiments devoted to computing, has led INFN in recent years to adopt the policy of a centralized review of all the computing need of such experiments through the CCR and the National Scientific Committees. This choice was made to avoid duplications, optimize spending and make sure that the requirements of the computing models chosen by the large LHC international collaborations (i.e., TIER-2 regional centres) would indeed be met. INFN also funds a national scientific computing centre (CNAF), which hosts the LHC TIER1 and presently meets the computing needs of several other experiments (i.e. BaBar, CDF2).

The proposed centre for scientific computing presented in this report will fit well in the context of computing within INFN that we have briefly outlined in this paragraph, while meeting the needs of local research activities.

2.4 Overview of LNF Research Activities

The Frascati National Laboratory is the largest structure of INFN, with about 350 staff, 50 temporary contracts, 140 employees of Universities guest scientists, and many external users. Details on the Frascati structure and activity can be found in [1,2]. The Laboratory is composed of three main units, i.e., Accelerator Division, Research Division and Administration.

The Accelerator Division runs several projects. The DAΦNE electron-positron storage ring which produces ϕ mesons at high rate, for the KLOE, FINUDA and SIDDHARTA experiments. A Beam Test Facility provides clean electron and positron beams from DAΦNE's linac. SPARC and SPARX, projects focused on construction and operation of free-electron lasers, and FLAME, a very intense laser, perform multidisciplinary research with emphasis on material science, biophysics and medical physics. The CNAO proton synchrotron, based in Pavia, designed and built partially in Frascati, will be a hadrotherapy facility for cancer treatment. The CTF3, CLIC, TTF, ILC, Super-B projects address issues in accelerator technology. Finally, DAΦNE produces synchrotron radiation light used by many experimental groups and external users.

Many major experiments in nuclear, sub-nuclear and astroparticle physics, and a strong theory group, are represented in the Research Division: KLOE, FINUDA, SIDDHARTA and NAUTILUS at Frascati, ALICE, ATLAS, CMS and LHCb at CERN, CDF at Fermilab, BaBar at SLAC, AIACE at JLAB, HERMES at DESY, GRAAL in Grenoble, OPERA and ICARUS at Gran Sasso, VIRGO at Cascina, WIZARD in space. The experimental groups in Frascati address basic research problems which range from the search for the missing block of the standard model, the Higgs boson, to the discovery of the new particles which could shed light on the Dark Matter enigma, to the physics of quark and lepton flavours, to the behaviour of quarks in nuclear matter and in atom-like structures, to the basic questions in theory and phenomenology.

Such a diverse and complex scenario is realized via a formidable effort in both design and construction of detectors and experimental apparata, and data analysis via scientific computing which employs state-of-art hardware and software techniques, such as computing farms and GRID technologies.

2.5 Computing Models

There are many possible ways of accomplishing the transformation of the raw data to a physics result using computers; these are different computing models. Typically, it is the scale of both (or either) the data set and the collaboration (number and geographical distribution of participating institutions, for instance) that make different models work best in a particular case; or it could be the specific policy for data access, or simply the kind of calculations which need to be performed (i.e., theoreticians).

The analysis of experimental data normally proceeds via a reconstruction stage, where raw data output from detector is transformed to physical quantities such as tracks, momenta, masses, and a physical analysis stage where physical quantities are combined to produce high-level information directly used for the discovery or measurement purposes of the experiment. Monte Carlo simulation is always used. In both cases, a very large number of interactions (”events”) are processed independently, and distributed models where a large number of computers physically even very far away and connected via the Internet analyze events independently are very commonly used.

On the other hand, theoretical and phenomenological computations very often use parallel or quasi-parallel algorithms carried out by computer farms concentrated in one site only.

The LHC experiments have all chosen computing models based on the use of the GRID where computing centers are made available around the world to sustain the data processing demands.These centers are distributed and configured in a tiered architecture that functions as a single coherent system. Each of the tier levels provides different resources and services.

The TIER0 is present only at the main site (for instance CERN but not only). It accepts RAW data from the Online Data Acquisition and Trigger System, and archives the packed RAW data to tape, distributing RAW data sets among the next tier stage resources (TIER1). The TIER0 also performs calibrations in order to get the constants needed to run the reconstruction tasks (RECO) and distributes the RECO datasets among TIER1 centers. The TIER0 does not provide analysis resources and only operates scheduled activities.

The second tier is the TIER1. There is a defined set of TIER1 sites, which are large centers in collaborating countries. These sites will in general be used for large-scale, centrally organized activities and can provide data to and receive data from all TIER2 sites. Each TIER1 center receives some subset of the datasets from the TIER0, provides tape archive of part of the RAW data and provides substantial CPU power for scheduled re-reconstruction, skimming, calibration, AOD extraction, and other data-intensive analysis tasks. The TIER1 stores an entire copy of the AOD while distributes RECOs, skims and AOD to the other TIER1 centers as well as to the associated group of TIER2 centers, and provides secure storage and redistribution for Monte Carlo events generated by the TIER2's (described below). A more numerous set of smaller TIER2 centers, with substantial CPU resources, provide capacity for user analysis, calibration studies, and Monte Carlo production. TIER2 centers provide limited disk space, and no tape archiving. TIER2 centers rely upon TIER1's for access to large datasets and for secure storage of the new data (generally Monte Carlo) produced at the TIER2. The Monte Carlo production in TIER2's will in general be centrally organized, with generated Monte Carlo samples being sent to an associated TIER1 site for distribution among the community.

In summary, the TIER2 sites provide: services for local communities, grid-based analysis for the whole experiment (that is, TIER2 resources are available to whole experiment through the GRID), and Monte Carlo simulation for the whole experiment.

Some research activities simply need some (or a lot) local computing power and storage, with local access. This is typically the case for theoreticians (CSN-4), or medium-sized groups with relatively large data samples (CSN-3) where processing tasks may still be faced using a large, but locally distributed computer cluster. In this approach, raw data are stored by the online systems in a temporary stage-in area usually made out of one or more RAID partitions using NAS or SAN technologies. The data is kept on disk for several hours for the online reconstruction to monitor the behavior of the experimental apparatus, then is automatically transferred to a Mass Storage System (tape) with a typical capacity of 1-2 PB. When massive data reconstruction is run, the files are extracted back from the tape, processed on a large Linux cluster, eventually filtered, and the final results are sent back into the MSS. Since tape operation take longer than disk operations, a cache area (disk) is used to hold a limited but significant fraction of raw o reconstructed data on disk for additional or customized online processing by the users. Such cluster structure is replicated without any particular coherence at all participating sites. The only difference is in size and robustness. In fact, at the main experimental sites, the cluster is larger and professionally managed. At smaller satellite sites, where only offline tasks are processed, the cluster is usually smaller and may be self-managed by the users. The computing models used in the accelerator physics is completely different from the previous one briefly described above. In this case there is the need to simulate the behavior of each single particle in the beam both to project new machines and to find the best working points. Several efforts are developing to reduce the computing time and increase the number of simulated particles, for example, some codes use the idea of macro-particle i.e. an agglomerate of many single particles that are treated as unique object, others treat part of the dynamics in an analytical way. In both cases there is the need to introduce several approximations that are in many cases too crude and anyway the number of the simulated particles is far from the reality to contain the computing time within the human limits. For these reasons the possibility of increasing the computing power becomes very important: this will allow a better simulation of the beam dynamics by increasing the number of the simulated particle and at the same time reducing the approximation used in the codes. It is also important to underline that most of the existing programs implement real parallel algorithms.

3 Review of Existing Computing Resources at LNF

In this section we review the existing computing resources and how they are organized and distributed.

3.1 LNF Computing and Networking Services (C&NS)

Presently the LNF run a centralized service which guarantees all the basic Intranet/Internet services such as network access and security, e-mail, web-mail, printing, authentication, web pages, centralized installations, DHCP and many other ancillary services which are needed by everyday operations. This group also provides a large support to several administrative services (i.e. Data-WEB, Servizio Informativo), used not only by the Laboratory but also the whole INFN. On the computing specific side, the C&NS provides a 5-node, 10-core cluster running Scientific Linux 3. This system is intended for general purpose scientific computing so it contains many of the common tools used by the nuclear and high energy physicists community plus a number commercial software tools to support more dedicated tasks.

All the computing hardware of the C&NS occupies an area of $\approx 100m^2$ situated in the ground floor of the Computing Centre building.

3.2 KLOE Computing Centre

The KLOE Computing Center is located in two different halls: Hall 1, at the third floor of the KLOE building, Hall 2 in building 14, close to the LNF Computing Centre. The two halls cover an area of approximately $180\ m^2$.

The centre is the only computing facility of the experiment. Using grid jargon it therefore acts as the TIER0-1-2 of KLOE, providing data storage and handling, computing power for data reconstruction, Monte Carlo production and physics analysis.

Massive data storage is provided by the use of two automated tape libraries, located in the two above mentioned halls. The first one with 12 drives and 5000 cartridges, the second one with 6 drives and 3500 cartridges. The total present capacity is $\sim$ 1 PB. Data storage on disk is also available for a total capacity of 100 TB, mainly used for DST staging.

Computing power is provided by a farm of 220 IBM PowerPC processors, for a total of about 140 kSI2k. During data taking about 50% of the power is used for the quasi-online data reconstruction, while the rest is left to Monte Carlo or physics analysis. Data analysis is performed using the KLOE AFS cell. Networking has been built with a LAN and SAN for data access using fiber channel for disks and tapes.

Since the year 2000, the system operates 24 hours a day 7 days a week, excluding short periods of programmed maintenance. It is run by a team of three people (two staff, one temporary contract) who are expert members of the experiment. The LNF Computing Centre only provides a link to the external world. Air conditioning and UPS are under the responsibility of the Laboratory.

3.3 ATLAS Proto-TIER2

The Laboratori Nazionali di Frascati host an ATLAS computing farm, that was one of the four Atlas TIER2's proposed for approval in 2005. After an in depth scrutiny, INFN approved three of the four centres, putting the Frascati Atlas farm in stand by as a "Proto-TIER2", pending the realization of the infra-structural developments needed for the installation of all the computing hardware foreseen for an a full size approved TIER2.

Despite its proto-TIER2 status, the Frascati farm participates fully to all ATLAS TIER2 activities as a member of the Italian ATLAS TIER2's federation. The ATLAS Computing Model makes substantial use of Grid Computing concepts, thereby allowing the same level of access to data and computing resources to all members of the ATLAS Collaboration [3], following the general model described in section 2.5.

The ATLAS off-line and analysis computing model is a hierarchical multi-tier model that consists of one TIER0 and 10 geographically distributed TIER1; each TIER1 is coupled with $3 - 4$ TIER2. A typical TIER2 is made by a set of nodes configured as servers or computing nodes, and a set of disk based Storage systems. The Grid middle-ware is suitably customized and regularly updated with the last INFN-Grid release. All services defined by the ATLAS-Grid community are installed and running. The ATLAS specific software is centrally managed by the experiment.

The ATLAS farm relies on the network and computing service infrastructure provided by the Laboratory's Network and Computing Service, which also provides support in hardware purchasing and commissioning, and in Grid middle-ware and Operating system upgrading activities.

The total amount of computing capacity, including the most recent acquisitions of 2008, is 170 kSI2k and 64 TB of raw disk (i.e. 6% the size of a typical ATLAS TIER2 in the year 2010).

This hardware is housed in one rack for the servers and computing nodes and another for the disks. Both racks are part of the LNF computing and network infrastructure, and both have local Ethernet switches that are attached to one 1 Gbps link core switch. Table 3 summarizes the available computing resources.

The proto-TIER2 manpower consist of 0.5 FTE from the C&NS for the manage-

Table 3: Summary of existing hardware resources in the Atlas Proto-TIER2 farm.

Nodes/Cores	kSI2k	Disk (TB)	OS	SW/LIB	MODEL
15/78	170	64	SL	gLite (INFN-GRID)	GRID

ment of operating system, Grid middle-ware and Storage systems; additional 2.0 FTE from the experiment are responsible for the experiment-specific issues like accounting management, production flow, farm monitoring, Federation and Grid meeting participation etc.

For urgent requests, the ATLAS personnel also takes care of system updates, always in cooperation with the Network and Computing Center. Moreover they interface the site with the INFN GRID community and the "Federazione Italiana TIER2 di ATLAS". A further 0.5 FTE is involved in a centralized Data Management activity for all the ATLAS community.

3.4 CSN-1 Other Activities

We summarize here the existing computing activities of other CSN-1 groups at LNF which have less impact on local centralized resources.

We start from the LHCB experiment, whose computing model[5] is similar to that of the other LHC experiments, with a CERN based major TIER1 centre, where the central production of data is performed, a series of 6 other TIER1 centres geographically distributed for reconstruction and user analysis, and a number of TIER2 regional centres. One important difference with respect to the other experiments, is the fact that the TIER2 regional centres will be exclusively devoted to Monte Carlo production. In Italy, the LHCb TIER1 and TIER2 centres are both hosted at CNAF, and all analysis activities of the Frascati group therefore are and will be done at CNAF and/or on personal workstations on site. The LHCb group does not envisage any major use of the SCS resources.

At LNF there is a research group involved in the R&D effort for the detector of the SUPER-B project. The group is involved in the study of the central tracking chamber and has taken the responsibility of the development of the simulation tools for the whole detector. This software activity has already started and is today mainly concentrated on the development of the source code; the work is carried out using personal workstations running Scientific Linux.

The NA62 group is presently involved in the construction of the apparatus, therefore there is no activity that requires any specific computing resources.

The CMS Collaboration searches for the Higgs boson and for new particles at the

CERN Large Hadron Collider (LHC). CMS utilizes a GRID-based, multi-tier computing model [6]. The CMS Frascati group plans to use local computing resources for physics analysis. The Frascati group is involved in the study of final states with di-muon pairs, profiting of the expertise of CMS muon detectors, in particular di-muon pairs from Z0, J/ψ and Υ decays. The group presently uses Frascati resources limitately to personal workstations, AFS disk areas for storage of small root datasets, and it relies heavily on the CERN computing infrastructure.

The CDF Collaboration studies top quark physics and searches for the Higgs boson at the Fermilab Tevatron Collider. The CDF Frascati group analysis interests cover topics in B and high-p_t physics. CDF reconstructs raw data and stores it on Fermilab-resident tapes. Each analysis group builds the ntuples of their interest using the CAF, a computing farm at Fermilab. For Monte Carlo simulation only, CDF uses non-Fermilab farms, such as the CNAF in Bologna(Italy), where sub-skims of interest for the Frascati analysis are also stored for easier accessibility. Personal workstations are used to perform local analysis on root ntuples.

3.5 CSN-2

The CSN-2 activities can be grouped in two main areas: ground based experiments and space based ones.

Ground based current astro-particle activities at LNF [2] include the search for gravitational waves using resonant bars (Nautilus/Explorer), neutrino physics underground (Opera, Icarus, BENE-INFN) and underwater (Nemo).

The only on-site experiment at LNF is Nautilus, which is run since a decade from the ROG Collaboration and, therefore, implements a well-established and longstanding computing model. The main task of the online computing is signal filtering aimed at the identification of burst-like (impulsive) events in a single detector. The current Nautilus data storage is of $\sim$2 TB/y including also simulated data and waveform catalogs for the astrophysical sources.

On the other hand, Opera at LNGS is an experiment based on nuclear emulsions and, therefore, implements a quite unusual computing model. The most challenging tasks are image processing and track reconstruction at the optical microscopes that scan nuclear emulsions. The microscopes (two of them are located at LNF) are equipped with dedicated graphic cards and front-end CPU's. Raw data are recorded in a local ORACLE database and filtered data are transferred to the Opera central database. The latter is located at LNGS and mirrored in Lyon (France). At present, OPERA does not make use of common computing infrastructure at LNF: emulsion data are stored in the microscope local machines and electronic detector data are processed in the online LNGS cluster. LNF, however, is currently responsible for the reconstruction of the electronic detector data and, therefore, also of the offline processing. Future development that can involve directly the LNF Scientific Computing Center are discussed in Section 4.2

Computing needs for other activities such as Nemo and Icarus are quite limited and do not require specific common resources.

Space based current LNF activities include two experimental areas: the consolidated astro-particle physics program of Wizard, culminated in the Pamela mission, and a newer effort concerning gravitation physics in the Earth-Moon system. The latter activity includes Lageos-Lares, MoonLIGHT, part of NASA's "Lunar Sortie Science Opportunity Program" and Magia, a candidate for an ASI "Small Mission" now in Phase A. These experiments expressed their interest in developing locally a joint computing resource for astro-particle physics and gravitation experiments in space.

Pamela's computing activities are shared among a main centre located at CNAF and the INFN Sections of Roma2, Trieste, Firenze, Napoli and Bari. Its software, data production and distribution are exclusive responsibility of INFN. Data reduction and simulations are largely done in a non-GRID environment, but there is a potential interest in migrating them to GRID.

The data storage system used is CASTOR, that is presently migrating to STORM. The data flow from Moscow is 5 TB/year. Currently, the Frascati group is accessing remotely the computing resources, and dedicates about 0.5 FTE to the software maintenance and installation.

The Pamela group is interested in starting up a local analysis activity in collaboration with the Tor Vergata groups of Pamela and of Solar Physics to study positron and anti-proton momentum spectra and correlation of PAMELA cosmic-ray spectra with important features and variations of the solar activity.

The data analysis of the gravitational experiments covers three topics: theoretical calculations; orbit reconstruction using the free data products of the International Laser Ranging Service and the calibration of the laser ranging data taken at the LNF SCF space facility. Part of the software involved in this work runs under Windows OS and part under Linux OS. The thermal and orbital analysis is now performed on a Dell Workstation with Dual Core 3GHz Xeon processors used both for code development and production.

3.6 CSN-3 - Nuclear Physics Computing Centre

In the last ten years, the needs for raw data processing of nuclear physics experiments operating at intermediate energies (several GeV) has entered a regime where hundreds of

TB raw data storage and large amounts of computing power are demanded.

The typical computing approach used to cope with this data production rate is a locally distributed computer cluster running Linux. On top of this, a batch submission system such as LSF or PBS is installed to allow massive job execution.

For new generation nuclear physics experiments such as ALICE at CERN and PANDA at GSI, the traditional approach of local distributed computing is clearly insufficient. Such experiments use a geographically distributed computing approach to cope with the storage and processing of multi-PB raw data sets. In fact, the PANDA experiment will produce roughly around 0.5 PB/yr, while ALICE will reach 2PB/yr.

The middle-ware that grants a transparent access to the worldwide distributed computing and storage resources composing the computing GRID of ALICE and PANDA has been developed by the ALICE collaboration and is knows as AliEn (ALIce Environment).

Non GRID-aware nuclear physics experiments (such as AIACE at JLAB or HERMES at DESY) run their offline analysis tasks on large clusters installed at the sites which host the accelerator machines. Smaller replicas or clones of the main cluster are run at the other collaborating sites, without any special coordination among them.

The operating system is almost always a Linux distribution, the most popular recently being Scientific Linux (SL) from SLAC and CERN. Among the most used packages and libraries used in the nuclear physics field we can list the ubiquitous tools such as ROOT, CERNLIB, CLHEP, different generators (PHYTIA, HIJING, FLUKA, GARFIELD...) and compilers.

The experiments named above constitute the large fraction of the computing power requests coming from the nuclear physics group at the LNF and a dedicated computing facility has been set up for this purpose. This facility is located into a $16m^2$ dedicated room in building 22. The room is air conditioned and electrically controlled/protected. A dedicated dual-pump heat removal system (400 kBTU/h) grants an adequate air temperature ($18C^{\circ}$) while a dual, redundant UPS system of 16 kVA fed by the LNF privileged electrical up-link grants a nearly 24/7 operation.

The files servers are based on the NAS technology and exports their volumes via NFS (UNIX) and CIFS (WINDOWS) for a net total of 20 TB. Part of the volumes are also backed up on LTO2 tapes using a node equipped with a 30 slot SCSI robotic library and a specific software (CA Brightstore ArcServ).

The computing cluster is composed by a 14 nodes running (slightly) different versions of Linux. The OS versions will be aligned when and if the computing model will require such operation. The nodes have different CPU architectures: the older machines run dual HT Xeon while the newer machines run four 3.1 GHz Dual-Core Xeon or two Quad-Core 1.86 GHz Xeon. The total CPU power is evaluated around 133 kSI2k.

Table 4: Summary of hardware resources of the CSN-3 group, located in the Building 22 computing centre.

Experiment	Nodes/Cores	CPU (kSI2k)	Disk (TB)	OS	SW/Lib	Model
AIACE	3/24	48.0	11	SL	Open SRC	Inter./Batch
HERMES	7/20	34.0	3	SL	Open SRC	Inter./Batch
ALICE	3/24	48.0	6	SL	Open SRC	GRID
PANDA	1/2	03.0	-	SL	Open SRC	GRID
Total	14/70	133.0	20			

Table 5: Summary of hardware resources in the FINUDA counting room.

Experiment	Nodes/Cores	CPU (kSI2k)	Disk (TB)	OS	SW/Lib	Model
FINUDA	6/52	120.0	25	SLC	Open SRC	Inter./Batch

FINUDA has been using the building 22 Linux cluster only for their Monte Carlo production. Their raw data storage, processing and offline analyses are instead performed on a 6-nodes, 44-cores custom cluster running Scientific Linux located into the FINUDA counting room. The total computing power of this system is evaluated around 120kSI2k. The FINUDA group also manages a backup/mass storage system based on one 32-slot LTO3, FC robotic library controlled via AMANDA.

Other experiments such GRAAL and SIDDHARTA run their offline analyses on dedicated high-end (Xeon) workstations administered by the personnel directly involved in the experiments while AMADEUS and VIP will entirely rely on the KLOE2 computing model.

In Tables 4 and 5 we summarize the available hardware resources and the software environments of the various experiments.

3.7 CSN-4 - Theory Group Computing Cluster

The research activity of the theory group is divided into two main areas: theory and phenomenology of high energy physics, theory and phenomenology of condensed matter and many-body systems. More details can be found in the Frascati activity report [2]. All of them need strong computing efforts to calculate quantities of interest to be compared with the experimental data coming from different experiments running inside and outside the Frascati National Laboratories. For example, the condensed matter group activity which is mainly linked to the Synchrotron Radiation experiments to provide the best theoretical framework for the interpretation of the data coming from different spectroscopies, typically needs the inversion of big matrices (more than 1000x1000, complex double precision) to calculate the whole energy spectrum of atomic clusters formed by hundred of atoms. This must be done for many different energies and several geometrical

Table 6: Summary of existing hardware resources of the CSN-4 computing cluster.

CPU/nodes	CPU (kSI2k)	OS	SW/LIB
14/28	29	Linux	Score

configurations of the atomic cluster. For these reasons, these applications work in a MPI environment and can be considered as *real* parallel applications. The theory group in the last several years has built and maintained a small cluster of three servers for a total power of ≈ 29 kSI2k. Table 6 summarizes these data.

These servers are supported by a data storage on disk for a total capacity of 1.2 TB. Everything are inserted in a rack located in a very small room in the high energy physic building provided with air conditioning and power supply. At the moment the condensed matter group is the major user of the whole cluster.

3.8 Accelerator Division

At the present time computing resources in the Accelerator Divisions are relatively limited, mostly due historical reasons and lack of man power to dedicate to the management of a centralized system. Most of the calculations for CNAO, CTF3, DAFNE and SPARC have been typically performed on large personal workstations (4-core) or, partially, on the C&NS cluster. Other more demanding ones have been performed on machines hosted by large computer centers (KEK, Berkley, UCLA). For example, to perform a scan in the machine parameters space to obtain a map of operating points for DAFNE, a 2 week run on 30 dedicated CPUs lent by the Tor Vergata University computer centre were used.

3.9 Summary

The result of our review is summarized in Table 7, which describe what computer resources are being used, and in Table 8.

We find that there is a limited amount of computing power available from the present LNF Computing Service, which is being used to some extent by people in the RD and the AD. KLOE has a large independent computing centre, managed by people belonging to the experiment. The LNF computing centre houses the Atlas Proto-TIER2, which is supported mostly by people from the experiment, but also receives support by the LNF C&NS. The nuclear physics community is using a relatively large computing cluster housed in Building 22, and managed by 2 people part time, which serves all the CSN-3 experimental groups present at LNF. Finally, the theory group is running a smaller

Table 7: Summary of existing computing resources at LNF dedicated to scientific computing. The CPU rating expressed in kSI2k are approximate; power consumption per CPU and TB may vary depending on the age or type of the hardware used. The KLOE hardware is located in two different buildings; the hardware in the computer centre hall is only disk storage.

	Nodes/Cores	CPU(kSI2k)	Disk (TB)	Tot Power (KW)	# racks
KLOE	37/184	140.0	100.0	40.0	
C&NS	6/22	15.0	54.0	8.0	1.0
Atlas Proto-TIER2	15/78	170.0	64.0	14.0	2.0
LARES	1/2	3.1	1.0	1.0	0.1
AIACE	3/24	48.0	11.0	3.0	0.4
HERMES	7/20	34.0	3.0	4.5	0.3
ALICE	3/24	48.0	6.0	2.7	0.3
PANDA	1/2	3.0	0.0	0.5	0.1
FINUDA	6/52	120.0	25.0	4.8	2.0
Theory Group	14/28	28.0	1.2	7.0	1.0
Total except KLOE	52/224	434.1	165.0	45.5	6.2

cluster of machines which are managed independently by 2 people working part of the time.

In total there are ≈ 2.8 FTE presently working on the maintenance of the various existing computing infrastructures, including 0.5 FTE, out of its total manpower of 6 FTE, dedicated by the C&NS to the support of the Atlas Proto-TIER2. The present level of support from the C&NS should continue at least for the initial startup period of the Scientific Computing Service, until dedicated manpower is devoted to it. The expertise of the people presently involved ranges from basic operating system and hardware support skills to high-level GRID software tools.

Table 8: Summary of existing computing resources at LNF dedicated to scientific computing. The manpower is divided in support to the group-specific software and hardware/software management of the systems.

	Total	Group SW	Sys. Manag.
KLOE	3.0		
Computing Service	0.5		0.5
ATLAS	2.5	2.0	0.5
LARES	1.0	0.75	0.25
AIACE	0.4	0.2	0.2
HERMES	0.1	0.05	0.05
ALICE	0.4	0.2	0.2
PANDA	0.1	0.05	0.05
FINUDA	1.5	1.0	0.5
Theory Group	0.5		0.5
Total except KLOE	7.0	4.25	2.75

4 Review of Computing Requirements of Research Activities at LNF

In this chapter we review the possible future evolution of the computing needs of the research groups at LNF.

First we have identified a number of activities which could immediately benefit from a prompt realization of the SCS, as the hardware resources could be immediately installed and put online in the new structure as soon as this would become available. We list them in Section 4.1.

Then we discuss in Section 4.2 how these activities could evolve over time and identified which new ones could need significant computing in the future and add to the existing ones.

To complete this exercise we picked a starting date to set the time scale. We chose July 1^{st}, 2009 as our working hypothesis, as it comes about twelve months after the publication of this study, a time we reckon sufficient for the implementation of the first phase of the plan we propose in Chapter 5, if accepted.

4.1 Immediate Needs

We present here all the outstanding requests for computing resources that could be fulfilled if the Scientific Computing Centre is approved and ready to start operations on July 1st, 2009.

4.1.1 ATLAS TIER2

The timely realization of the infrastructure could allow the full deployment of the LNF ATLAS TIER2, that could be dimensioned, e.g., like the last approved ATLAS TIER2 in Milan. With the new purchases programmed in the second half of 2008, the Milan TIER2 will provide almost 100 TB of storage disk and 289 kSI2kof CPU, which is about double the actual disk space and CPU power of LNF proto-TIER2, but still less than half of ATLAS requests for a TIER2 in 2008.

For what the internal network infrastructure is concerned, the ATLAS TIER2 are presently upgrading their internal network to 10 Gbps connections between racks containing computing nodes and storage servers.

4.1.2 CSN-3 Cluster

The Nuclear Physics Group cluster can be transferred from Building 22 to the new SCS structure without any particular problem. Besides the obvious move of the storage and computing elements (three 42U-racks filled at 70% capacity), the transfer may also include network switches, the racks themselves and UPS units, which may be usefully re-employed during the start-up phase. The CSN-3 group is interested in establishing synergies with other computing farms which have similar computing models and which will be transferred into the SCS room as well. The exact extent of the overlap will be defined after the start-up phase. Moreover, the nuclear cluster will have the chance to scale up in a cleaner way, smoothly migrating to more efficient configurations (for instance, leaving NAS storage in favor of SAN). The only component which may be retained into the previous location in building 22 is the backup system, without any significant impact on the new planned setup, since the backup server is network based and does not have to reside physically close the storage servers. This service is very specific and run by dedicated hardware and personnel.

The Finuda offline system can be easily contained in a single rack. Since this system is also used for some critical online tasks, its transfer to the new SCS room is not convenient for the Finuda group. Nevertheless, if the present offline computing capacity should double in the event of a third Finuda data taking run, the installation of new worker nodes and disk servers may be done in the new infrastructure.

4.1.3 CSN-4 Cluster

The realization of the infrastructure could allow the accommodation of the existing cluster in the new building without any problems. At the same time the theory group could use the financial support asked to the national INFN CSN-4 to build a cluster formed by 11 dual core servers, which will increase the group's computing capability to satisfy the immediate needs for the software developments of the Condensed Matter group and for starting the non-perturbative QCD numerical studies of QCD by lattice simulations.

4.1.4 CSN-2 Space Physics

As soon as the SCS will become available, the joint space physics analysis at LNF (Pamela data and LARES/LAGEOS orbits) could immediately begin. The resources needed for this program amount to one quad core server running Linux and 5 TB of storage capacity. Work on Pamela will start on satellite data already available. LARES work will start first on LAGEOS data, available since 1976, for which a total of about 1 TB of storage will be needed, including control samples from other satellites. Concerning the initial LARES data rate in 2009, there are no official estimates from the ILRS at this time.

For the LARES-LAGEOS work on thermal backgrounds, modeling of laser ranging and theoretical predictions, one quad-core server running Windows OS and 2TB of storage are needed.

For the thermal, orbital and spin analysis, there is an urgent need to separate development on the existing workstation from batch production on the requested server.

These resources will be used also for space technology experiments of CSN5, Alt-Criss and ETRUSCO, in which the same PAMELA and LARES groups are involved, respectively.

4.2 Evolution of the Existing Research Activities

We describe here what are the possible future development of computing needs beyond the startup phase of the SCS.

4.2.1 KLOE-2

An experiment to test a new interaction scheme for DAFNE is currently under way. The goal is to increase the luminosity of the machine by a factor of at least 3. Under this hypothesis, a proposal for the continuation of the KLOE physics program with an upgraded detector has been presented to the Laboratory (KLOE-2). Operations can start as early as Spring 2009 and can last up to year 2012-2013. The total amount of data (real or simulated) will be a factor 10 larger than that handled at present for KLOE.

KLOE-2 will make use of the previously described KLOE computing model (see). A detailed plan for upgrading the system at the required level has already been elaborated. The needed technology is already available on the market, for both storage and computing. There is no need of new space, since the new hardware can be accommodated in the two existing KLOE computing halls. However cooling and UPS power must be upgraded (see discussion of the infrastructures in Section.5).

4.2.2 ATLAS TIER2

According to ATLAS computing plans, in 2010 ATLAS TIER2's are expected to have reached a long term configuration with 800 TB of disk space and 1500 kSI2k of CPU power.

The ATLAS TDR foresees an important increase of resources for 2012, almost a factor 1.4 both for computing power and storage capacity. It's difficult now to evaluate the total need of electrical power and conditioning capacity, required in 2012. Old nodes will be replaced by future developed apparatus that clearly will concentrate more disk capacity or computing power in the single units, requiring less electrical and conditioning resources.

Taking into account that the previsions contained in this document made references to the present electrical requirements of the apparatus we do not expect to require more than the given specifications.

4.2.3 CSN-1 Other Activities

In the next two years the Super-B group will be involved in the R&D work finalized to the writing of the Technical Design Report of a detector for the Super-B accelerator. This work has already started with the study of the central tracking chamber and the Monte Carlo simulation software. As soon as the initial phase of software development is complete, the group could use the available resources provided from the SCS to generate Monte Carlo events to study detector effects and analyze physics channels relevant to the definition of the detector performances. A preliminary estimate of the resources which would be required for this activity is the equivalent of 8 CPU cores and 2 TB of disk space.

The reconstruction of the data for the NA62 experiment will be done at CERN. The local group could benefit from 1 server quad core with AFS and 8 TB disk storage dedicated to data analysis. With these resources the use of the CASTOR tape system at

CERN by the local group could be minimized and the analysis greatly sped up.

CMS utilizes a GRID-based, multi-tier computing model , [6]. The CMS Frascati group plans to use the resources of the scientific computing service for on site physical analysis activities (event selection studies, fitting, generation of small samples of Monte Carlo specific signal events, etc). For this purpose, a reasonable preliminary estimate would be 70 kSI2k of CPU and 20 TB of disk storage. For this work, CMS requests system support of common use libraries in analogy to ATLAS. The manpower needed is very small.

CDF reconstructs raw data and stored them on Fermilab-resident tapes. Each group builds the ntuples of interest using the CAF, a computing farm at Fermilab. For Monte Carlo simulation only, CDF uses non-Fermilab farms such as the CNAF, where sub-skims of interest for the Frascati analysis are also stored for easier accessibility. Personal workstations perform local analysis on root ntuples. The availability of a serious support from the Frascati Scientific Computing Center will allow to move locally skimming and filtering, as well as a better efficiency in Monte Carlo simulation. CDF Frascati requests infrastructures to host computing and storage with size 40 kSI2k of CPU power and 8 TB of disk space, for a total one-rack physical space.

4.2.4 CSN-2 Ground Based

The gravitational wave analyses performed in resonant bars have been mainly carried out seeking for burst-like sources. The computing challenges increase substantially when searching for non-impulsive sources and they are mainly related to the determination of Fast-Fourier-Transforms for long periods of time. An estimate made by the ROG collaboration of the requested computing power for a non-impulsive analysis of the resonant bar data is of about 100 kSI2k and, therefore, could be accomplished using resources available in a Scientific Computing Center.

For what concerns Opera, common computing resources could be exploited effectively for data and Monte Carlo processing downstream the LNGS/Lyon central database, aimed at the production of ROOT files with high-level data descriptions (OPERA Data Format).

The production storage (Monte Carlo and Data) assuming 200 days of data taking at nominal CNGS intensity would not exceed 3.5 TB/y and the CPU power needed is of the order of 30 kSI2k. Therefore, in the medium term (2009-2013) OPERA-LNF considers with interest the opportunity to have a computing centre managing a cluster aligned with CERN for what concerns operating system, batch systems and maintenance of the most common HEP libraries.

Finally, it is worth mentioning that in the framework of BENE-INFN it has been considered the possibility of exploiting laser-plasma acceleration techniques for non-conventional neutrino sources. These studies were based on particle-in-cell (pic) simulations done in JAERI-Kansai (Japan). Dedicated pic simulations at LNF could be done sinergically with other laser-plasma acceleration research programmes related to the exploitation of FLAME.

4.2.5 CSN-2 Space Based

Space physics analysis at LNF is motivated by the availability of Pamela data and by the new space test facility, the Satellite Laser Ranging Characterization Facility (SCF), to be expanded to become a Space Laboratory inside a clean room.

The total manpower of space experiments is about 10 FTE. In the next three years the group plans to exploit the full range of analysis topics in the field of astro-particle physics and gravitation physics in the Earth-Moon system. Computing resources will be used also for other CSN-5 space activities: SCF characterization of laser retro-reflector payloads of Galileo and GPS (Etrusco experiment); BTF calibration of astro-particle payloads for the International Space Station (Alt-Criss experiment). These planned activities translate into the following additional needs: one Quad Core Xeon Server running Linux OS, with 10 TB of storage for astro-particle and gravitation data analysis, and one Quad Core Xeon Server running Windows OS, with 10 TB of storage for experimental calibration and for thermal background analysis.

Longer term requests will be assessed in 2010, when the analysis work will be consolidated.

4.2.6 CSN-3 Cluster

The dimension of the Nuclear Physics cluster will most likely double in the next couple of years. This infrastructure will indeed follow the upgrade path foreseen for those experiments which, having ended their data taking runs, will enter their final analysis phase. Moreover, the local Alice group which is fully committed into the electromagnetic calorimeter Monte Carlo simulations, will also need increasing their computing resources up to ≈200 kSI2k of computing power and 150 TB of storage. A significant contribution may come also from Panda which has already performed a GRID data challenge run, and will need more resources for Monte Carlo simulations. A reasonable figure in this case is ≈100 kSI2k and 10 TB of disk space. Considering the very good shape of the 12 GeV upgrade plans for J-Lab, also the Aiace experiment has significant figures: roughly 50 kSI2k and 10 TB of disk space. Finuda may have to face another season of data taking with the RUN III and the dimension of its installation will have to double, which will mean additional 120 kSI2k and 25 TB of disk space.

4.2.7 CSN-4 Cluster

The theory group will be forced to increase his computing capability in the near future because new theoretical activities, based on non-perturbative numerical studies of QCD by lattice simulations, are starting now in connection with the new experiments carried out at LHC and inside the Frascati Laboratory. They will need an increase of the computing power and data storage of at least a factor of 5 to be completely fulfilled. For this reason, the upgrade foreseen at the startup of the SCS will have to be followed by another bigger one.

It must be noted that this request is in agreement with the policy of the national Scientific Committee that is already supporting similar initiatives within different INFN sections and laboratories. The new cluster cannot be accommodated in the small room actually used by the theory group, for this reason it become crucial to move our servers in the new building, provided with the right supporting tools, where all clusters devoted to the scientific calculations will be located. The new cluster must save the MPI environment to still run real parallel applications, and, at the same time, it could become member of the Virtual Organization named Theophys to increase the capability for farm-computing applications.

4.2.8 Accelerator Division

In the near future, the largest need for CPU power for the accelerator group will come from Dafne, Sparc-Lab (Sparc, PlasmonX, Flame, Beats,...), and SparX projects.

For DAFNE the main requirement in terms of CPU power comes from the need of beam-beam simulation and electromagnetic beam dynamics simulation: both electrons and positrons beams must be simulated with their interactions and behaviour after hundreds millions of turns. This type of studies are now crucial to find the best working point of the machine.

For both Sparc-Lab and SparX, the computational needs come from beam dynamics and plasma simulations. For example, a complete simulation required by the PlasmonX group needs a computing power of the order of 500 kSI2k. Many computation of the Accelerator Division could be performed using the resources that will be made available at the SCS. Often these calculations use proprietary software programs that need special installations and are subject to licenses. The SCS could take over the management of license acquisition and distribution. The centralization of these tasks optimize the use of

Table 9: Summary of computing resources available at the startup of SCS, on July1 1^{st}, 2009.

	CPU(kSI2k)	Disk (TB)	Power (KW)
ATLAS Proto-TIER2	289.0	100.0	24.0
CSN-3 Cluster	133.0	20.0	23.0
CSN-4 Cluster	72.0	1.2	13.5
CSN-2	12.0	8.0	1.5
Grand Total	495.0	129.2	62.0

Table 10: Computing resources required in 3 years from today.

	CPU(kSI2k)	Disk (TB)	Power (KW)
CSN-1	1640.0	828.0	75.6
CSN-2	205.0	35.0	4.2
CSN-3	911.0	261.0	31.4
CSN-4	380.0	6.0	13.5
TOTAL	3136.0	1130.0	129.2

human and financial resources.

Moreover, additional ad-hoc resources could be added and be easily hosted in the new structure, as the Accelerator Division requirements grow in time.

4.2.9 Synchrotron Radiation Facility

No special computational requirements come from the synchrotron radiation light facility users: typically the amount of acquired data during experiments is small and users will bring their data at home for the analysis using network and/or few Dads.

4.3 Summary

We can summarize the results of our survey of the future requirements in three tables, which represent the expected amount of resources at the startup of the SCS set on July 1^{st}, 2009 (Table 9), in three years from now (Table 10), and in five years from now (Table 11). We have assumed that the Atlas group will get in 2009 the same level of financial support received this year, that the CSN-3 cluster will be transferred as is and that the theory group will be granted the funding for a first upgrade of the present cluster. The CSN-2 resources are those required by the space physics activities, which could be acquired in early 2009. The numbers summarized in Table 10 assume that in 2009 the Atlas TIER2 is

Table 11: Computing resources required in 5 years from today.

	CPU(kSI2k)	Disk (TB)	Power (KW)
CSN-1	2240.0	1148.0	109.1
CSN-2	205.0	35.0	4.8
CSN-3	1460.0	530.0	59.7
CSN-4	380.0	6.0	13.5
TOTAL	4285.0	1719.0	186.9

approved in full, and therefore will scale up gradually in size to equal the other approved centres over the following two years. The resources reported are those requested by the Atlas computing TDR for the year 2010, which in our working example coincides with the solar year 2011; we think this is consistent with the delayed approval of the Frascati site. It is also assumed that the Alice and Panda computing activities is increased and funded to scale up resources from the current levels. We consider also the eventuality that the theory group receive enough funds to be able to scale up their CPU resources by a factor of 5. The numbers summarized in Table 11 take into account a significant ($\approx 40\%$) in the Atlas TIER2 resources and the CSN-3 cluster reaching its final configuration; both these assumptions are of course subject to verification, and may not become true. On the other end, we assume no further increase in the resources in use by CSN-2 and CSN-4 activities as there are no reliable estimates now. It is worth pointing out that all the power consumption number are computed based on today's technology which most likely will be subject to improve in the next five years in the direction of less power consumption per unit computing power; similar considerations hold for the disk capacities and densities.

All in all, the two highest SCS potential users appear to be the CSN-1 and CSN-3 groups, which are dominated by the Atlas TIER2 and the Alice and Panda computing clusters respectively.

While the CSN-1 numbers include the requirements of CMS and CDF which are of some relevance, we also recorded a relatively large number of minor requests coming from other smaller experiments in all the CSNs. Given that these represent a small fraction ($\approx 10\%$) of the total computing power of the SCS, they can easily be made available to all users, in the form of interactive and batch access, using the already deployed hardware thus realizing the synergy that the SCS will bring about.

5 Scientific Computing Service

The goal of the Scientific Computing Service (SCS) is to provide, at the same time, computing support for research activities requiring geographically distributed computing (GRID-aware) and local distributed computing.

From the survey performed on site described in Chapter 3, we have identified and characterized a number of already existing “dedicated” computer resources.

Many groups run similar or largely compatible systems which would greatly benefit just by sitting side by side into a common framework; the gain in terms of optimization of running and infrastructure costs is clear and could be realized even at the SCS start-up.

But the optimization will be even greater when the participating systems will have to be scaled up to cope with the increased computing needs expected by the evolution of the computing activity of the various research groups. Besides the obvious sharing of infra-structural costs, and scale economies in purchasing components, the de-facto compatibility among the computing models observed in many cases could lead to a partial sharing of the resources.

We consider different phases in the deployment of the SCS; a start-up phase in which the most immediate needs can be accommodated with limited manpower and cost and a subsequent phase in which the system can be scaled up to accommodate the computing needs of existing or possible future activities in a time scale of the next 3 to 5 years.

5.1 Deployment Plan

We discuss here a possible evolution with time of the SCS. It is anticipated that at the startup the different groups participating to the project will transfer their existing resources into the new infrastructure. At that time, the overall computing power of the system will just be the arithmetic and incoherent sum of the participating subsystems.

As we have mentioned in the previous sections, we chose July 1^{st}, 2009 as the starting date. We recommend to deploy all needed infrastructures in the SCS hall before the first racks are put in; this means fitting the SCS hall with all the plumbing and wiring needed by power and cooling systems which can cope with the final expected loads. In this way these systems can be up-scaled with time as the computing resources (thus power consumption) increase by just changing elements like power supplies, UPS, transformers, fan-coils, etc., without performing any major civil engineering work in the hall.

We estimate that about 1 year will be necessary from the time the decision to go ahead is taken and the first operation of the facility; this would include the realization of the final engineering project, the bidding process and the actual realization of the infrastructures.

Assuming that the decision is taken no later than July 2008, this sets the start time on July 1st, 2009.

5.1.1 Startup Phase: July 1st 2009

At the startup date there will be a number of users ready to take advantage of the new facility. These are the computing systems that are already operating elsewhere in the LNF: the CSN-3 and Theory clusters and the Atlas Proto-TIER2. The computing power and the corresponding electrical and cooling power needed are summarized in table 9.

This table also shows the projected increase of requirements in the first 2 years, driven by the increase in the TIER2resources, which will only happens if INFN approves a fully fledged Atlas TIER2in Frascati.

It is reasonable to assume that in the first year of this phase there will be manpower support by the same people who have been supporting the individual systems and experiments until then. This means that there will be about 2.8 FTE (see Table 8) working together to transfer all the machines from their current physical location to the SCS hall.

We think that in this phase there should also be a computing technician to begin develop the expertise and know how about the systems operating in the SCS, to provide long term support and take over some of the load from the experiments' personnel as the SCS plateau to a smoother standard operating regime.

This number of people will be necessary and sufficient for the job.

5.1.2 Phase 2: 2012-2014

In a subsequent phase, between 3 and 5 years from the start, there could be more computing resources at LNF for the Alice and Panda groups. In particular the Alice group may get INFN approval for the funding of a significant amount of computing power locally, as we report in Tables 10 and 11. Similarly for the Panda group.

If the Atlas TIER2is approved during the startup phase, then it would grow and reach its full operational dimension in 2011, slowly increasing afterward.

It is important to stress that if not all the resources shall eventually be approved and funded by INFN, the accompanying infra-structural cost will be limited to the money spent in the predisposition of the SCS hall.

As the deployed computing resources augment, so should the manpower. When the Atlas TIER2, if approved, reaches its full size and operational regime, and before the Alice and Panda systems reach their full size, a second Computing technician should be devoted to the SCS. One additional technician would be needed later on, if the SCS reaches its full size.

5.2 Required External Support

5.2.1 Distributed File System

The SCS relies completely on the support of the AFS lnf.infn.it cell provided by the C&NS. In the cell will reside the home directories of the SCS users.

5.2.2 Network and Security

The SCS relies completely on the support of the security and network infrastructure provided by the C&NS, which will provide to the SCS machines secure access to the wide area network.

5.2.3 Backup System

The new SCS infrastructure will mainly host production data and despite the thousand of TB of storage capacity, a full structured backup solution is not foreseen.

The existing LNF Network Service already provides a fully backed-up AFS area. This service is supported by two robotic libraries controlled via Tivoli Storage Manager.

The “new” user areas of the SCS will be attached to the existing AFS tree or may be backed up installing a Tivoli client on that specific sub-cluster. The policy is to backup only the relevant information such as source code and documents, which do not pose a significant extra load on the existing backup system. The raw, generated or reconstructed data will not be backed up since it is assumed that a full disaster is not likely to happen (all disk partitions will be RAID5 or higher with Hot Spares); moreover, this kind of data is usually reproducible, if at the cost of some time.

5.3 Infrastructures

5.3.1 Requirements

The document “Service Level Description between ROCs and sites” [4] formalizes the services that a site provides to its Regional Operation Centre, and vice versa. The document is still in preparation but some points are already well defined: the site must provide a minimum hardware configuration and a given set of services; the availability must be higher than 95% of solar time, a percentage which is measured by routine checks centrally executed; software updates installation must be executed within the maximum time

Table 12: Summary of the evolution with time of the estimated power requirements of the SCS operation, expressed in KW.

Year	1/7/2009	1/7/2010	1/7/2011	2012-2014
Total Power	62.0	94.1	129.2	186.9
Extra Load on C&NS hall	53.6	85.7	120.8	178.5
Extra Load on LNF	12.0	48.1	83.2	140.9

agreed with the ROC; the site must provide at least one system administrator who is reachable during service hours and the site must respond to *trouble tickets* within one working day.

These requirements make a sufficient level of redundancy in computing and network apparata, infrastructures and manpower essential.

The support of all other activities required by the other research activities is by far less stringent that the one required by a TIER2as far as level of redundancy, up-time, response to user requests etc. Therefore any service that guarantees the support at TIER2level is more than adequate for all other activities.

5.3.2 Infrastructures

To determine the characteristics of the infrastructure needed to fulfill the requirements described above, we have summarized in Table 12 the most important driving parameter, which is the total power. It is worth noticing that the power required at the start up of the SCS obviously includes a large fraction ($\approx 85\%$) which is already supplied (paid) by the Laboratory as it serves hardware already installed and running. From the perspective of the load on the C&NS building power and cooling systems however, the situation is different as in this case only a small fraction ($\approx 15\%$) is already deployed in the hall, the rest being located elsewhere as explained before. Together with the total power we therefore give in Table 12 also the extra load at the C&NS and the extra load on the Laboratory as a whole.

Based on this data, we propose the solution outlined in the following sections.

5.3.3 Site

The most favorable location for the SCS is the ground floor of the present Computing Centre. Despite the fact that presently the ground floor of the building houses (some unused) offices, the original project foresaw the destination of the entire building’s ground floor to be the housing of the computing hardware (CPU, disks, switches, etc.).

The floor in that area, for example, is an elevated floor with some of the cooling and power services already installed under it. The same is true for the sprinkler fire system.

We give here two sketches of the Computer Center ground floor in its present configuration (fig.1, top drawing) and a possible solution to house the Scientific Computing Service (fig.1, bottom drawing), which can be implemented by rearranging the existing movable walls. In the proposed solution a hall with an area of $\approx 100\ m^2$ can accommodate up to ≈ 30 racks.

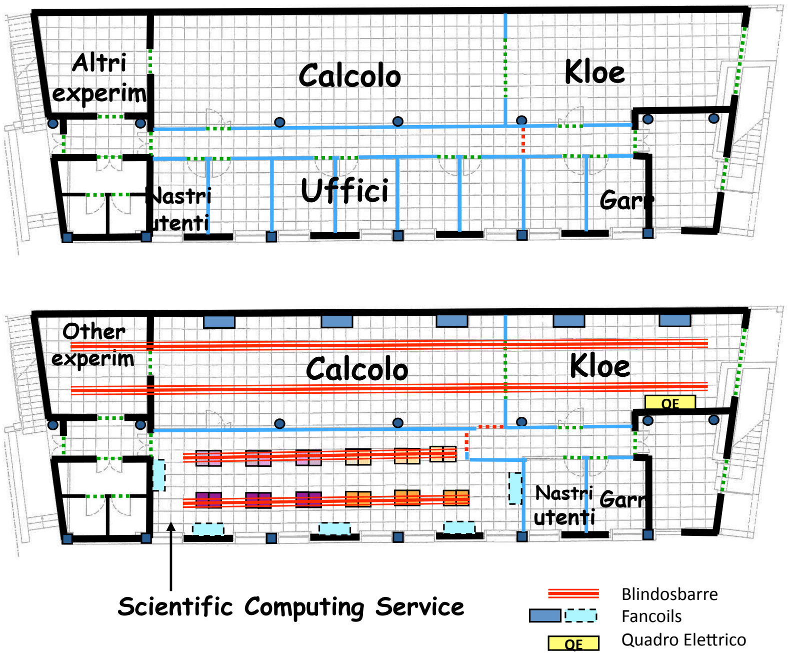

Figure 1: Present layout of the ground floor of the LNF Computer Centre (top). Possible layout of the Scientific Computing Centre (bottom).

5.3.4 *Cooling Power*

The SCS hall occupation will be characterized by low mean power density, so a traditional configuration is proposed, with water cooled air conditioners sending air flow under the floor. The air distribution should be balanced by accurate floor grids positioning, and a dedicated computer simulation should be performed to avoid hot spots.

The cooling power could be delivered by the near DAFNE chilled water system, as for the existing Computing Centre. At the moment, enough cooling power is available to fulfill SCS needs, but the reliability and continuity of the service has to be improved for the whole Computing Centre.

Among the other upgrades, a backup cooling system is foreseen to be realized for the sensible users. The installation will be staged as much as possible to follow the evolution of the SCS; the pumping station, piping, electrical switchboard and connections should be completed within the first step. Then the cooling backup system and part of the air terminals will be added, as well as the control and supervision updated.

The preliminary design will aim to guarantee the right approach to the reliability and the availability requirements of the service, according to the INFN-CCR Guidelines [8] for the computing center plants. Among the others, it will give strong prescriptions on the commissioning, on the control system and on the maintenability of the plant. The preliminary design of the plant could be done by LNF engineers, and a detailed project by a engineering firm is needed for the final bid. The total cost estimate, including 20% contingency, is 150 K€ for the initial stage and 90 K€ for the second phase.

5.3.5 Electrical Power

A redundant power distribution system, which allows to supply electrical power to critical loads via a twin source UPS and a standard circuit, is at the moment available in the Computing Centre building. However, with the present hardware only a fraction of the total envisaged load can be powered. For this reason, an upgrade has been studied to supply the additional power demand for the new computing devices of SCS that will gradually reach 180 KW and provide a redundant power supply for the back-up cooling system in order to reduce the dependability from the main system.

The study has taken into consideration as much as possible the reutilization of components where appropriate, and a staging of the deployment which follows the development of the computing centre with time.

The electrical infra-structural work will be split in two tasks:

- Upgrade of the substation switchboard and of the building switchboard, with the extension of the internal power distribution system. This job has to be scheduled at the beginning and will deliver up to 100 KW worth of computing devices.

- Installation of a second UPS and a second insulation transformer to reach the full power demand.

In the first stage all the infra-structural work in the computers hall will be completed, so that all the subsequent upgrade stages will be performed outside the hall, thus minimizing the disruption of computing service. The total cost estimate is, including 20% contingency, 60 K€ for the initial stage and 90 K€ for the second phase.

5.3.6 Network Access

The network infrastructure is made by a set of local switches, one per rack, interconnecting all the computing nodes installed in the rack. Moreover a backbone switch interconnects all the local switches and provides the connection with the computing centre. At present, network connections are based on 1 Gbps channels, either through dedicated switches or using the general purpose ones already installed in the different buildings where the computing resources reside now. At the SCS startup, local switches need to be installed where necessary and some of the existing one must be replaced by switches supporting 10 Gbps ports, in particular where a more performing access to the storage is foreseen.

Table 14 includes a possible scenario for the cost of the network infrastructure needs in the different SCS phases. The cost is referred to the switches that should be considered as part of the infrastructure provided by the Laboratory.

5.3.7 Safety

The envisaged use of the present hall in the computing centre hall does not present any critical aspect with respect to safety and/or fire hazard. The existing extinguishing system already in place in the building will be sufficient to cover the additional hardware installation foreseen by the present study. For this reason, no significant additional cost is considered necessary.

5.4 Manpower Resources

While in the initial phase of the deployment of the SCS one technician will be sufficient, as he/she would work in conjunction with the people who have been so far involved in running the existing facilities, we estimate that if and when the Atlas TIER2is approved and ramped up to its full operating size, to run the SCS a total of 2 IT technicians will be needed, with expertise in the following:

- Linux OS
- Management of local and distributed network architectures
- Job scheduling systems in distributed environments
- Computing and storage resource monitoring systems, locals and distributed
- Remote managing and OS configuration in a highly distributed environment

- Scripting languages (shell scripts, Perl, etc.)
- Management and backup of large data sets
- Management of users programs and libraries
- Personal computer clusters architectures

One additional IT technician, with the same profile, dedicated full time to the SCS may eventually be needed if and when the SCS should grow to its full size.

The SCS should be led by a software engineer who takes on the responsibility of the whole service. She/He should keep the relations with the various users and research groups, provide guidance to the technicians in the group, and provide a liaison with the Computing and Network Service.

5.5 Operating Costs

We did not evaluate in details what are the running costs of the SCS, but we give here an estimate of the major contributing factors. The most important is the energy consumption for electric power supply and cooling, which is evaluated at 1.5 K€ per KW per year including both. Another cost is the maintenance of the computing hardware, which however is limited to the hardware which is bought with Laboratory funds, as the hardware acquired by the experiments would come with its own maintenance contracts. Then there is the cost of the maintenance of the electrical and cooling systems, which is a small perturbation of the present operating costs of the existing systems.

Table 13 shows a summary of only the dominating costs over the years; manpower cost is not included. The last row of the table show what is the cost after subtracting the present level of power consumption already paid for by the Laboratory. Given that presently the electric power bill for the Laboratory runs at ≈ 1.3 M€/year, the extra cost due to the SCS operation is small in the first two years and raises slowly up to $\approx 15\%$ of the total electric bill at its very maximum capacity.

Table 13: Summary of the estimated dominating operating costs of the SCS in K€.

	1/7/2009	1/7/2010	1/7/2011	2012-2014
Total electric power	93.0	141.0	194.0	280.5
Extra charge on LNF operating funds	18.0	72.0	125.0	211.0

5.6 Funding and Manpower Profile

To the deployment scheme described in Section 5.1 correspond the funding profile of Table 14 and the manpower listed in Table 15. The initial cost covers all the work needed in the Computer Centre hall, while the cost in 2011 covers the needed upgrades when the installed power exceeds $100KW$, which will only happen then. A fraction of this money could actually be spent in 2010 in order to smooth out the profile.

Table 14: SCS funding profile in K€.

	1/7/2009	1/7/2010	1/7/2011	2012-2014
Cooling System	150.0	0.0	110.0	0.0
Electrical System	60.0	0.0	90.0	0.0
Computing Hardware (Networking)	45.0	0.0	20.0	0.0
Total Infrastructure Upgrades	255.0	0.0	220.0	0.0
Operating Extra Cost	18.0	72.0	125.0	211.0
Grand Total	273.0	72.0	345.0	211.0

It should also be noted that part of the cost of 2009 includes upgrades to the systems that will benefit not only the SCS but also KLOE-2 and the C&NS; in fact, some of this money would have to be spent in any case to allow the upgrade of the KLOE Computing Centre infrastructures in preparation for a KLOE-2 high luminosity run. The operating cost listed in the Table 14 is only the extra cost with respect to the operating cost of the present computing hardware, as explained in Section 5.5, it therefore represent the actual investment of fresh money for the Laboratory.

Table 15: SCS manpower profile.

	1/7/2009	1/7/2010	1/7/2011	2012-2014
IT Technician (FTE)	1.0	1.0	2.0	3.0

Acknowledgments

We would like to thank M. Pistoni and S. Vescovi for their important contributions to this report, and P. Santangelo for useful discussions.

References

[1] M. Benfatto et al. (Eds), Future activities at LNF - Working group report, Frascati preprint LNF-05/33(IR).

[2] "LNF 2006 Annual Report", M. Antonelli ed., LNF-07/10 (IR).

[3] The ATLAS Collaboration, ATLAS Detector and Physics Performance TDR, http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/TDR/access.html.

[4] Ioannis Liabotis, John Shade, Service Level Description between ROCs and sites, https://edms.cern.ch/document/860386/0.5.

[5] "LHCb Computing Model", LHCb note, 16 Dec 2004, CERN-LHCb-2004-119.

[6] CMS Computing Technical Design Report, preprint LHCC-2005-023.

[7] P. Astone *et al.* [IGEC-2 Collaboration], Phys. Rev. D **76** (2007) 102001.

[8] "Linee guida per gli impianti dei centri di calcolo INFN - Impianti di condizionamento dell'aria", Luigi Pellegrino, nota CCR-05/01.