

Summary of the Plan for ensuring the Infrastructure needed to meet the Computing Needs for Physics at CERN in the Next Decade

Ian Bird, Tony Cass, Bernd Panzer-Steindel, Les Robertson – CERN/IT

Executive Summary

Power is now and for the foreseeable future a major concern for any organisation that requires to increase its computing capacity. At CERN the current estimates of the requirements of the LHC experiments suggest that the power available in the Computing Centre will be insufficient by the beginning of 2010. It is therefore urgent to extend the power and cooling infrastructure for computing at CERN while preparing for the longer term requirements that will arise from the projected upgrades to the accelerator and the detectors.

Within a fixed budget line for physics computing at CERN, the cost of providing the additional infrastructure for power and cooling will be a major fraction of the total expenditure for the foreseeable future. We estimate that this will limit the growth rate of the computing capacity to about 30% per year. This contrasts sharply with the 100% per year growth rate that we have experienced over the past 18 years, and we must question whether this will satisfy the basic needs. Nevertheless, even this very limited growth rate will require twice the current available electrical power by 2015.

This is also a problem for the regional centres, which are assumed to provide 80% of the overall computing resources for LHC, and several of our partners are also planning or already building major extensions of their computing infrastructure (e.g. FermiLab, STFC/RAL, IN2P3, SLAC). Despite this, however, the overall regional centre plans at present for 2012 fall 25% or more short of fulfilling the experiments' requirements. It is therefore unrealistic to look to them to *increase* their share of the overall load. CERN must continue to provide its share of the capacity that will be required for the lifetime of the LHC accelerator.

Preliminary work began over one year ago to explore options for addressing this problem¹. Unfortunately it was not possible to include in this study the appropriate engineering experts from TS Department due to the higher priority of the LHC installation work, but from the beginning of December 2007 the study group has been extended to include these experts.

The findings of the preliminary work undertaken during 2007 are included in this note. Four options have been studied: extending the capacity of the current Computer Centre; locating CERN computing equipment in an external computing (hosting) facility; purchasing a full service from an off-site supplier; constructing a new facility at CERN. Extension of the current Computer Centre is not possible due to lack of sufficient spare power on the Meyrin site, and purchasing a full external service is ruled out for both financial and practical reasons.

Constructing a new facility at CERN is significantly less expensive than using an external hosting service, and so this is the primary option retained. We must therefore proceed rapidly now with an in-depth market survey and technical study to prepare a proposal for the design and construction of a new facility on the CERN site, using the most energy-efficient technologies available. The new facility is unlikely to be available before the end of 2009 and so it is necessary also to study the options for renting space in a nearby hosting centre to be used as a stop-gap solution to meet the requirements for LHC.

¹ [Evolution of Power and Cooling Requirements for CERN Central Computing Services](#) - A.Cass, B.Panzer, L.Robertson – CERN/IT - 10 November 2006

1. Background, Resource Requirements and Long Term Evolution

The four LHC experiments have each developed a plan defining their anticipated needs for computing resources at CERN and in the external regional computing centres to the end of 2012. The annual increase in data volumes and a gradual increase in the sophistication of the reconstruction and analysis software will drive a steady increase in the offline computing demands of the LHC experiments. These [plans](#) are based on assumptions of accelerator performance, program efficiency, the behaviour of their computing models and the reliability and efficiency of the distributed computing service. These assumptions will be verified only when we have had the experience of a full year of data taking and analysis, but we should note that estimates for all previous experiments and accelerators have been under rather than over estimated.

Beyond 2012 we can only speculate on how the computing requirements will grow. With increased luminosity and upgrades to the accelerator and detectors the requirements are likely to grow at an even accelerated rate. For example, with increasing luminosity the number of overlaying minimum-bias events will grow. For example the 50 nsec option considered for SLHC would lead to a fivefold increase of the overlaying events from the normal number of 20 at design luminosity. In order to keep the occupancy in the tracking detectors at a manageable level, the experiments plan for new tracking detectors with smaller cell-size and consequently increased number of channels. The increased complexity of events will require a significant additional demand in computing resources for reconstruction and physics analysis. We must also foresee the requirements for detector and accelerator design and the possibility there may be new experiments as we look out towards 2020.

It is clear therefore that we must be prepared for a continuing increase in the computing capacity that will be required at CERN to support the physics programme. However, one of the factors that will limit computing capacity growth rates in the coming years is the electrical power needs. This is discussed in later sections of this note, but there is one major implication that must be considered here: the cost of this infrastructure will be significant and in the first instance we must assume that it will have to be fully funded from the physics computing budget.

The evolution of requirements beyond 2012 shown in Figure 1 takes some account of the considerations outlined above, but constrains the total cost within a physics budget capped at its level at the end of the current Medium Term Plan (2012). This results in an average annual rate of increase of CPU and disk of only 30% after 2012, with the magnetic tape capacity increasing only linearly at 19 PetaBytes each year.

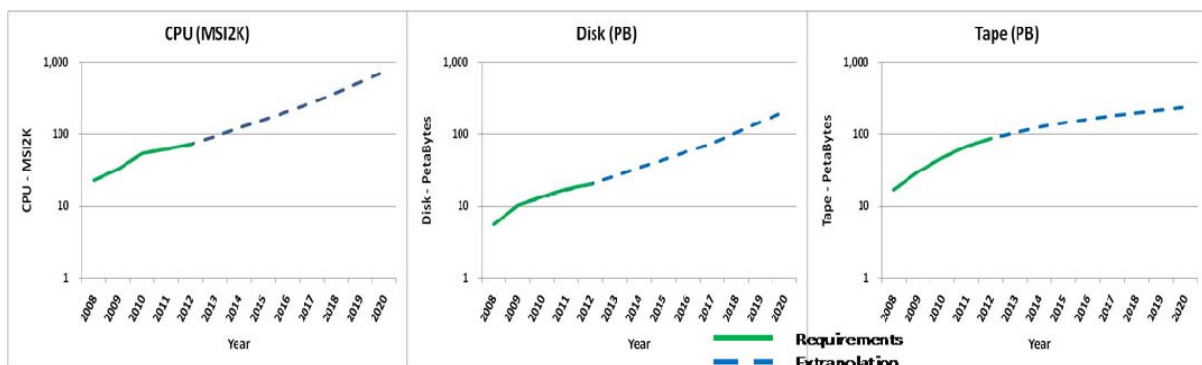


Figure 1 - Evolution of planned computing capacity at CERN

This power constrained growth rate looks very low compared with the experience of the past eighteen years, as can be seen from Figure 2, which shows the evolution of computing capacity at CERN since the installation of the first computer in 1958. The rate of increase during the period from 1990, when we started using single chip processors, averaged about 100% per year. The 30% per year rate growth projected for the LHC era is

similar to that which was experienced during the period of stagnation of the mainframe years, when computer cycle rationing forced experiments to look for alternative solutions such as building their own computers.

While we take the requirements profile of Figure 1 for planning purposes, it must be recognised that this will only support a basic physics programme. All proposals for accelerator and detector upgrades as well as for any new experiments should include an analysis of the computing capacity requirements and the funding to provide them.

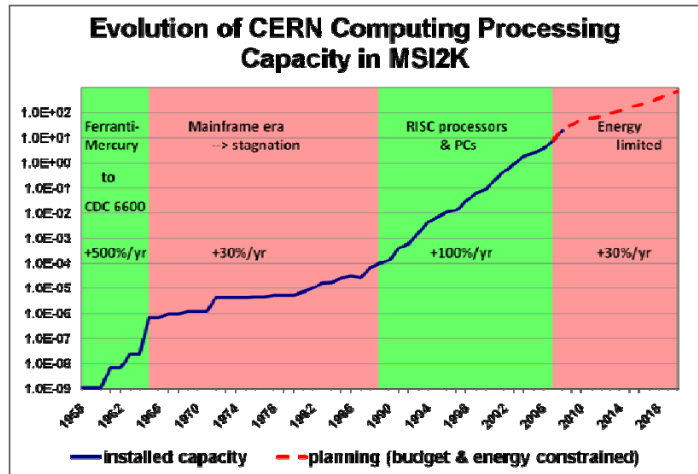


Figure 2 - Evolution of Computational Capacity at CERN from the First Computer to the Present Day

2. Power Considerations

For the foreseeable future the power requirements of computer processors are expected to continue to grow with the computational capacity. We expect to see some improvements in the technology, but also in the packaging and system integration that will slowly increase the power efficiency (capacity/watt), but an improvement rate of 30% per year (which is our assumption) is at the optimistic end of the scale. Figure 3 shows the evolution of these parameters over the past five years for the single board dual processor systems on which most of HEP computing has been based. The improvements since 2004 are largely due to the introduction of dual and then 4-core processors. However, as far as the basic technology is concerned there are at present no solutions proposed that will bring the capacity/power growth rate back to the levels experienced during the 1990s. The technology issues and assumptions concerning the evolution of power requirements are discussed in more detail in a recent IT report².

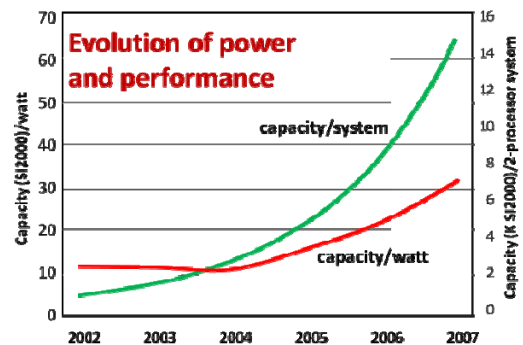


Figure 3 - 5-year evolution of system power and performance

HEP labs are of course by no means the only organisations to be hit by this problem, and several industrial providers of large scale computing services (e.g. Google, Amazon, Microsoft Networks, and Yahoo) are building new facilities in areas with low energy costs. At the same time work is going on to develop more energy-efficient approaches to computer centre design and construction.

Figure 4 shows the power requirements estimated to be needed at CERN for the LHC physics resources shown in Figure 1, along with that needed for the current non-LHC experiments and the general computing services. It does not include the

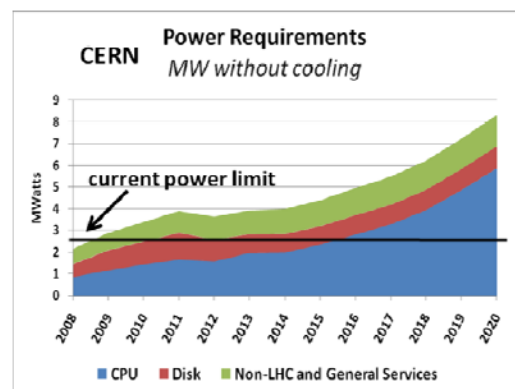


Figure 4 - Power Requirements - Input Power (without cooling)

² [Technology and cost trends – CERN T0 and CAF computing Facility](#), Bernd Panzer-Steindel CERN/IT, 12 December 2007

power required to cool the equipment. The CERN Computer Centre has at present a maximum electrical power capacity for computing equipment of 2.5 MW, with additional power to cover distribution losses and cooling. The anticipated load will exceed the current capacity by the end of 2009, depending on the performance of the accelerator during its first year of operation. While estimating the longer term requirements is necessarily imprecise, decisions on the computing equipment to be used and the acquisition process need only be taken a year before the capacity is required, enabling the equipment installed to be adapted to the evolving needs of the experiments. This is not the case for the infrastructure for power distribution and cooling, where the civil and electrical engineering work required has longer timescales. It is therefore urgent to plan for a substantial increase in capacity to be available by the beginning of 2010. The options that are being considered are outlined in later sections of this note.

3. Options

Several options for fulfilling the energy needs for the next 10-15 years have been or are being investigated:

- Additional power and cooling capacity in the current Computer Centre.
- Construction of a new energy-efficient computing facility on the CERN site.
- Renting space in an off-site computer hosting facility.
- Purchasing off-site computing capacity.

Each of these options is discussed below.

3.1 Additional Power and Cooling in the CERN Computer Centre

An upgrade of the CERN Computer Centre was initiated as part of the first phase of the WLCG project in 2002 and carried out progressively during the subsequent years. The upgrade was financially constrained to fit within the WLCG project budget, which itself was not fully funded at that time. In addition, it was generally expected in 2001 that the computer industry would be able to find an effective solution to the power problems by the end of the decade. The upgrade was designed using the standard approach to cooling common at that time, with a power density of 2 KW per square metre. The upgrade was also (for financial reasons) constrained to fit within the existing building (B 513 - constructed in 1971 to house mainframes and supercomputers of that era) without any interruption to the operation of the computing services and minimising the civil engineering work inside the building. Adding additional power and cooling capacity within this building while it is operating as the main CERN computer centre operating 365 days a year is impractical. In any case no additional power can be made available on the Meyrin site before 2013 and so this option is not being pursued.

3.2 A New Computing Facility on the Prévessin Site

Preliminary work has been undertaken to investigate the feasibility and costs of constructing a new energy efficient computing facility on the Prévessin site, where sufficient power and a suitable site are available. An alternative of using an existing building (B.904) could also be considered. In order to estimate the total cost of a purpose built facility we use a cost estimate provide by one of the **openlab** partners, Intel, of \$6/watt for constructing a computer centre with a power capacity of several MW, including civil engineering, power and cooling infrastructure, with full diesel backup. Similar estimates have been obtained from other sources, and also estimates that are significantly lower and higher. A full-scale market survey and technical study including the assistance of experts in the field of data centre construction would be required to produce accurate estimates. We would envisage a construction providing initially 2.5 MW increasing to 5 MW as required, and with the possibility of further extending the building at a later stage. The total cost of the 5 MW project is estimated at about \$30M (CHF 33M). The facility would be expected to have a lifetime of about 20 years, but

the budget model referred to in section 4 would finance the construction in stages over a period of up to 4 years. For comparison with other options with an annual rental charge, we amortise the construction cost over 10 years and use CHF 0.05/kWh as the CERN cost for electricity. This gives an annual cost of CHF 1.1M per MW per year.

3.3 Renting Space in a Computer Hosting Facility

In this option CERN purchases equipment and installs it in rented space in an off-site computer centre. The hosting facility provides power distribution, fully backed up with diesel generators, cooling, physical connectivity to network connections rented by CERN, and a basic level of operation. CERN is responsible for installing the hardware and all software, monitoring the service and for intervening in the event of problems, system re-configuration, etc. It is therefore only practical for a moderate-sized installation if this is in the Geneva area. A more distant installation would require CERN-managed staff to be available at the remote site, which would only be justified in the case of a major installation.

Using data obtained from two local hosting companies we estimate the annual cost of renting space to provide 1 MW of electrical power with 40 Gbps of network connectivity between the co-location site and CERN as CHF 3.6M. This may be compared with the cost of a new facility on the CERN site of CHF 1.1M per MW year (see section 3.2).

The conclusion is that, although the overall cost is significantly higher than the cost of providing the service on-site, it is an option to be considered seriously as a short term measure if it is not possible to provide additional on-site capacity before the needs of the LHC experiments exceed the current capability.

A more detailed discussion is provided in [Appendix 2](#).

3.4 Renting Compute Power via the *Computing Cloud Model*

The *cloud* computing paradigm is not a new concept; it has been developed and promoted already by companies like HP and IBM with names like *utility computing* or *ubiquitous computing*. Other keywords used were *on-demand-computing*, *pay-as-you-go IT services* or even *grid-computing*. There are many thousands of companies offering web-hosting services including E-commerce applications, either on dedicated servers or as part of a virtualised installation. Purchasing a 1 MSI2000 computing resource for 12 month from the larger companies would cost about CHF 500K and offload CERN by about 25 kW. These calculations are based on list prices from the corresponding web sites. A competitive tender would presumably bring somewhat lower prices. In addition CERN would have to pay for the network connectivity. There is a wide (upward) spread of prices - up to a factor of 5 more expensive.

Another type of cloud computing service is that offered by Amazon. Their service is not based on renting physical servers, but rather on time slices and computing slices. Purchasing 1 MSI2000 from Amazon for 12 months would cost about CHF 2.3M. Specific details about their computer centres and their network connectivity to the Internet are sparse.

Cloud computing services would not be suitable for the complex and critical services (such as physics databases) that could be re-located off-site in the Geneva area, nor would they be suitable for storing large amounts of permanent data. On the other hand they could be used to offload activities where data could be sent with the processing task and any output recuperated at the end of the job. An example of such Tier-0 activity is RAW data reconstruction, and CAF examples would include some calibration and ESD/AOD processing tasks - if the appropriate network performance would be possible. In 2009 we could off-load 5-10 MSI2000 for such applications.

A prerequisite for any cloud service would be a detailed investigation of operational issues including monitoring, remote intervention, operating system environment (Linux versions, virtual machines), ability to install grid services (workload management system, etc.).

The following table compares the annual costs of the three solutions considered for 40 MSI2000 of compute capacity requiring 1 MW of electrical power in 2009.

< ----- annual cost in MCHF ----- >				
Comparison of the cost of different solutions providing 40 MSI2000 capacity for one year (equivalent to 1 MW-year in 2009)	New facility at CERN	Local hosting company	Dedicated cloud	Amazon cloud
Building, infrastructure, electricity, networking	1.1	3.6	20.0	92.0
Equipment cost (purchased in 2009, amortised over the 3-year lifetime, including installation, network, etc.)	2.7	2.7	-	-
Operations by CERN staff	0.6	0.6	-	-
Total annual cost	4.4	6.9	20.0	92.0

In conclusion, cloud computing is not suitable for a sufficiently wide range of applications to enable CERN to off-load more than a few hundred kW of power. In addition, even with deep discounts, it is clearly not a financially viable solution and will not be considered further.

A more detailed discussion is provided in [Appendix 3](#).

4. Costs

Figure 5 shows the estimated costs for fulfilling the computing requirements of the LHC experiments outlined in Figure 1 taking account of the power requirements as discussed in section 2. The parameters used for estimating cost evolution of the basic computing technology are discussed in the IT report referred to in section 2 on page 2. The cost of computing centre infrastructure has been estimated using data provided by one of the **openlab** industrial partners as explained in section 3.2. The graphic also indicates the budget available for physics services in the Medium Term Plan. The largest single cost line is the allowance for the major investments which will be required at CERN for additional infrastructure to fulfil the expected power profile of Figure 4.

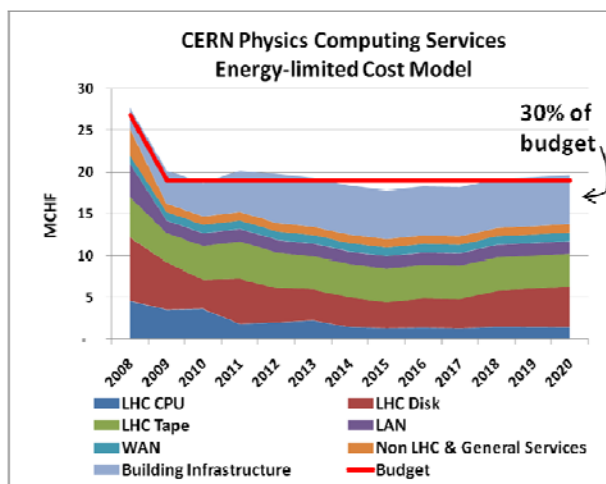


Figure 5 - Cost Estimates for Physics Computing Services

This problem is also being experienced by regional centres. At present the regional centre plans for 2012 fall 25% or more short of fulfilling the experiments’ requirements (see [Appendix 1](#)). The regional centres will do well to improve on this and match the expectations of the LHC experiments and so CERN should not look there for relief from its commitments.

5. Strategy

The conclusion from the above discussion is that an on-site solution at CERN will be the least expensive solution for increasing the computing capacity at CERN, but that a short-term co-location option could be considered while waiting for a new on-site facility to be prepared. Cloud computing services should not be further investigated at this time.

The proposed strategy outlined below aims to ensure that we have sufficient computing capacity available for the needs of the LHC experiments and the general site computing services for the next ten years. The project should foresee a later expansion of the facility to be operational in the second half of the next decade, but decisions on this should not be taken until 2011/2012 when the requirements will be better understood in the light of practical experience with LHC and advances in computing technology and computer room energy engineering will be available.

Preliminary work has begun on planning the design and construction of a new computing facility on the Prévessin site, with a minimum electrical power for the computing equipment of 2.5 MW capable of expanding to 5 MW, fully backed up by UPS and diesel generators. The infrastructure of the building should be designed to use the most power-efficient approaches to cooling available today, for example using for as long a period of the year as possible direct cooling by external air. The layout within the building should use a modular approach, allowing an extension of the infrastructure at a later date to fulfil the full computing requirements through to the end of 2020.

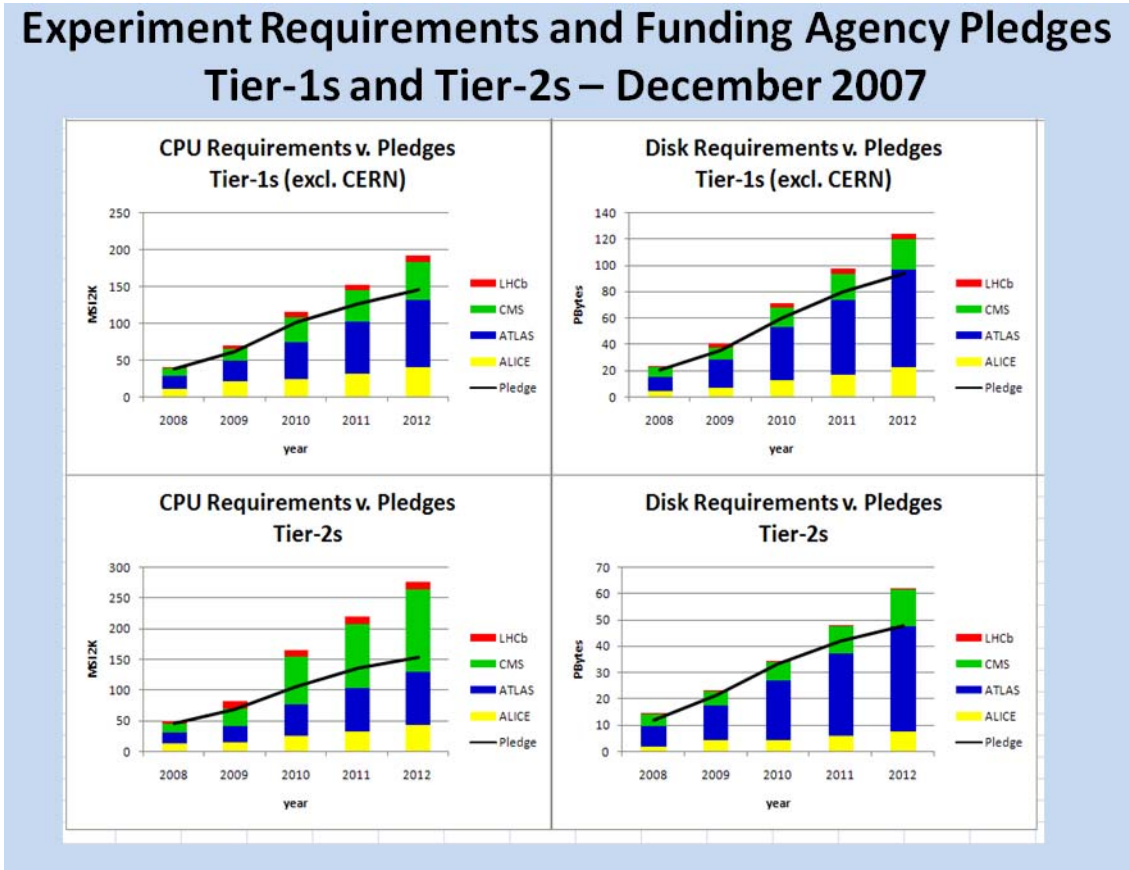
The requirement is to have the first phase in operation by the end of 2009, when the current computing centre on the Meyrin site will have run out of power. It is however unrealistic, given the delays inherent in CERN's acquisition process, to have a new facility on-line within two years and so a more realistic goal is to have the new centre operational by the end of 2010. This would require a stop-gap solution to be found for 2010 if the accelerator start-up is on schedule.

The first stage is an in-depth technical study and market survey to establish the best way to proceed, using as far as possible external experts in computer centre design and construction. At this stage the possibility of awarding a single contract for the design and construction should be retained, as well as other options such as using independent design and construction contracts. The investigation will be done in close collaboration between IT and TS experts. IT will provide a staff post for two years to compensate TS for the time of their civil engineering, electrical distribution and cooling and ventilation experts. IT has appointed a project manager for this stage of the project – Wayne Salter.

The first milestone of the preliminary project is to prepare a design proposal by the middle of the year, including estimated costs and timescales, and a financing plan, sufficient to enable a decision by the management to be taken.

In parallel with the preparation of the design proposal further study should be made of options to provide the required capacity during 2009/2010 within the power limitations of the existing computer centre. This should include: the use of co-location facilities in the Geneva area; identifying temporary solutions on the CERN sites; early replacement of installed equipment to take advantage of improvements in power efficiency.

Appendix 1 – Evolution of Experiment Requirements and Funding Agency Pledges in Tier-1s and Tier-2s – December 2007



Appendix 2 – Renting Space in a Computer Hosting Facility

In this option CERN purchases equipment and installs it in rented space in an off-site computer centre. The hosting facility provides power distribution, fully backed up with diesel generators, cooling, physical connectivity to network connections rented by CERN, and a basic level of operation. CERN is responsible for installing the hardware and all software, monitoring the service and for intervening in the event of problems. It is therefore only practical if this is in the Geneva area.

There are no fundamental network constraints preventing the relocation of IT services to a local hosting company. Two independently routed fibre connections can provide the redundancy and bandwidth (40 GB/s or more may be required) necessary for remotely located systems to function as effectively as if located in the CERN Computer Centre. Provided the space is located within a reasonable distance, initial estimates from Services Industriels de Genève (SIG) suggest that the necessary dark fibres could be rented for 7 KCHF/month; an initial setup up cost of around 100 KCHF would be required to cover equipment costs at each end and SIG installation charges.

Informal contacts with two hosting companies in the Geneva area, SafeHost and GigaPlex, suggest that space capable of supporting an equipment load of 2 kW/m² could be available for rent at a cost of around 300 CHF/m²/month. It has not been possible to obtain a quote for space capable of supporting heat loads above 2 kW/m². Since almost all services that could be relocated to a hosting company can generate heat loads above 2 kW/m², the space requirement, and thus the cost, will be independent of the actual space requirements (i.e. rack layout), but rather be determined solely by the total power demand. Electricity consumption is not included in the space rental cost; actual usage is charged at a slight premium above SIG rates or at about 0.20 CHF/kWH currently.

Apart from the tape robotics, there is no requirement in terms of day-to-day normal operations for most of the IT services to be located in the CERN Computer Centre, as the appropriate remote monitoring and software maintenance facilities are already in place. An advantage of this type of solution is that all capacity is supplied with diesel-backed power, and so it would be suitable for housing any services with a limited requirement for network bandwidth between the off-site facility and the CERN site. The major concern for equipment at a remote location relates to service interventions by vendors in the event of hardware failure. Here, there are concerns concerning the current contracts and practices, which require the presence of CERN staff during interventions and specify CERN as the location. These issues can presumably be addressed in future contracts, and be negotiated for existing contracts if the location is close to CERN.

The following table shows the annual cost of using such a facility to provide 1 MW of electrical power with 40 Gbps of network connectivity between the co-location site and CERN. The table does not include the cost of the equipment or operation which would be incurred whether the equipment is installed at CERN or in a local off-site facility. This may be compared with the cost of a new facility on the CERN site of CHF 1.1M per year (see section 0).

Requirements		Cost (kCHF)			
kW	Area	Space	Power	Network	Total
1,000	500	1.800	1,750	80	3,630
<i>assumes rental rate of 300 CHF/m²/month and electricity cost of 0.20 CHF/kWH</i>					

The conclusion is that rental of space in a nearby hosting company is feasible as a short term measure if demand for power, either critical or overall, exceeds the capacity available in B513. Although the overall cost is significantly higher than the cost of providing this service on-site it is an option to be considered seriously as a short term measure if it is not possible to provide additional on-site capacity before the needs of the LHC experiments exceed the current capability.

Appendix 3 – Renting Compute Power via the *Computing Cloud Model*

The cloud computing paradigm is not a new concept; it has been developed and promoted already by companies like HP and IBM with names like *utility computing* or *ubiquitous computing*. Other keywords used were *on-demand-computing*, *pay-as-you-go IT services* or even *grid-computing*. There are many thousand companies offering a web-hosting service including E-commerce applications (~2,500 in Germany alone). Some of them are offering an additional service, which allows the renting of a whole server (so called root-server packages) on a monthly basis including a certain amount of free data traffic (measured in Terabytes per month). The Linux servers used to provide these services are closely monitored by the company with a 24/7 service and all equipment is on a UPS system with diesel backup.

The second biggest web-hosting company in Europe is STRATO (www.strato.de) with two large computer centres in Germany (30,000 servers, full UPS+Diesel power, 35 Gbps connection to an internet exchange point in Frankfurt, 1.5 million customers). Two other companies of similar size are also based in Germany (www.hosteurope.de www.server4you.de) and have currently a similar service and price offering.

Purchasing a 1 MSI2000 computing resource for 12 month from these companies would cost about CHF 500K and offload CERN by about 25 kW. These calculations are based on list prices from the corresponding web sites. A competitive tender would presumably bring somewhat lower prices. In addition CERN would have to pay for the network connectivity to the Frankfurt internet exchange point.

There are many more companies offering these services (www.ultraserver.co.uk, www.eurovps.com, www.nine.ch, www.coolhousing.net etc.), with a wide (upward) spread of prices - up to a factor of 5 more expensive.

Another company offering a cloud computing service is Amazon. Their service is not based on renting physical servers, but rather on time slices and computing slices. Purchasing 1 MSI2000 from Amazon for 12 months would cost about CHF 2.3M. Specific details about their computer centres and their network connectivity to the Internet are sparse.

Cloud computing services would not be suitable for the complex and critical services (such as physics databases) that could be re-located off-site in the Geneva area, nor would they be suitable for storing large amounts of permanent data. On the other hand they could be used to offload activities where data could be sent with the processing task and any output recuperated at the end of the job. An example of such Tier-0 activity is RAW data reconstruction, and CAF examples would include some calibration and ESD/AOD processing tasks - if the appropriate network performance would be possible. In 2009 we could off-load 5-10 MSI2000 for such applications.

A prerequisite for any cloud service would be a detailed investigation of operational issues including monitoring, remote intervention, operating system environment (Linux versions, virtual machines), ability to install grid services (workload management system, etc.).

For comparison with other solutions we note that 40 MSI2000 of compute capacity from a dedicated cloud facility (off-loading 1 MW of electrical power) would cost CHF 20M per year, and from an Amazon cloud would cost over CHF 90M per year. The equivalent annual cost of doing this at CERN is estimated at around CHF 5M as shown in the following table.

In conclusion, cloud computing is not suitable for a sufficiently wide range of applications to enable CERN to off-load more than a few hundred kW of power. In addition, even with deep discounts, it is clearly not a financially viable solution at present.