CERN-RRB-2010-071

Computing Resources Scrutiny Group

C.Bozzi (Italy), T. Cass (CERN), C.Diaconu (France), D.Espriu (Spain, *Chairman*), J.Flynn (UK), M.Gasthuber (Germany), D.Groep (The Netherlands), A.Lazzarini (USA), W.Trischuk (Canada), B.Vinter (Nordic Grid), H.Renshall (CERN/IT, *Scientific Secretary*)

INTRODUCTION

The Computing Resources Scrutiny Group (CRSG) is an independent committee established by the WLCG Memorandum of Understanding whose members are selected by the Funding Agencies represented in the C-RRB. The purpose of the CRSG is to inform the decisions of the Computing Resources Review Board (C-RRB) for the LHC experiments.

According to the WLCG MoU, every year the CRSG shall scrutinize

- The resource accounting figures for the preceding year
- The use the experiments made of these resources
- The overall request for resources for every experiment for the following year and forecasts for the subsequent two years
- The CRSG shall also examine the match between the refereed requests and the pledges from the institutions
- The CRSG shall make recommendations concerning apparent under-funding.

This report summarizes the deliberations of the CRSG regarding the usage of the computing resources by the four LHC experiments during 2010. We include some items related to the CRSG work program for 2011 and about the composition of the CRSG itself.

A first comprehensive usage report was presented in the April 2010 C-RRB summarizing the activity of the four experiments and the WLCG collaboration during 2009 For the first time there was real data to be taken and analyzed, even if the total amount was much lower than originally hoped. In the elapsed months of 2010 the LHC has increased its luminosity by several orders of magnitude, the experiments' computing model have been subject to some stress in preparation of the summer conferences and the Tier 1 and Tier 2 centers integrated in the WLCG collaboration have been asked to deal with substantial amounts of real data. The new scenario requires us to provide to the C-RRB, according to the CRSG mandate, a detailed account of the usage that the experiments have made of the computing resources and of the availability of the pledged resources.

The CRSG has not received up to now, except in the case of the ALICE collaboration, any estimate of the resource request anticipated for 2012 and 2013, and consequently we have nothing to report to the C-RRB on this respect.

Recognizing the special characteristics of the present combined run, extending over two years, the uncertainties that remain in the current computing models and the need for redundancy/contingency, the CRSG committed itself in the previous C-RRB (April 2010) to undertake a revision of the 2011 scrutiny ahead of the October 2010 C-RRB in the light of the usage of the resources up to that moment. This revised scrutiny has not been deemed necessary as the experiments seem to have enough resources to fulfill their data taking and analysis and, on the other hand, a review of the usage of the WLCG resources reveals an increasing optimization of their use. In addition no modifications of the scheduled running for 2011 are expected.

The LHC schedule

The expected schedule of the LHC is an essential ingredient of the scrutiny even though a fraction of the computing needs do not scale with running time.

After the Chamonix meeting, in February 2010, it was agreed that, barring unforeseen circumstances, the LHC should run for the best part of 2010 and 2011 (the estimate is 8 months per year), with only a relatively short break at the end of 2010 and beginning of 2011. It is expected to accumulate an integrated luminosity of 1 inverse fb. The energy has been fixed at 7 TeV (3.5 GeV + 3.5 GeV) . About 10% of the time is expected to be dedicated to heavy ion (HI) physics. After this long run a long shutdown would follow to enable the machine to reach the design energy in 2013 and the following years.

For the scrutiny the relevant quantity is the total number of seconds when the beam is declared to be stable and good for physics. After a number of conversations with the CERN management it was decided to adopt the following scheme:

Live time: 30 days/month = **720** hours
Folding in efficiencies 720 x 0.7 x 0.4 = 201.6 effective hours/month = **725760** sec/month

RRB year	RRB year start	RRB year end	Months (max) Data taking	Total live time (in Ms)	Pp	HI
2009	**LHC start**	**May '10**	**3**	**2.2**	**2.2**	**0**
2010	**June '10**	**March '11**	**8**	**5.8**	**5.1**	**0.7**
2011	**April '11**	**March '12**	**8**	**5.8**	**5.1**	**0.7**
2012	**April '12**	**March '13**	**0**	**0**	**0**	**0**
TOTAL			**19**	**13.8**	**12.4**	**1.4**

For the elapsed months of 2010 the real scheme has been as shown in the following table. Note that 3.5 TeV beams started on March 30th 2010. We shall not discuss here the period with 950 GeV and 1.18 TeV beams that were too short and irregular to draw useful statistical conclusions even though these periods were crucial for commissioning.

Year	Start	End	Months data taking	Total live time (in Ms)	Pp	HI
2010	***April '10***	***August '10***	**5**	**2.3**	**2.3**	**0**

Comparing with the theoretical expectations table we see that, on average for the indicated period LHC delivered approximately 460000 seconds per month compared to the 725000 seconds theoretically expected (63%), which represents an excellent achievement in the first year for a machine of this complexity. What is even more encouraging is that in the last 10 weeks the machine is accumulating data taking hours at a pace that is only slightly below the theoretical target.

The number of events reflects the reduced live time. On average the experiments have recorded 40% fewer events than originally envisaged. However, for the first time the experiments are dealing with volumes of data comparable to the expectations. This represents, with the possible exception of ALICE, a first real test for the computing models and for the WLCG as a whole.

Interactions with the experiments

The recommendations of the previous C-RRB report in April 2010 urged the experimental collaborations to submit a detailed account of the usage of the resources by September 1st. In the meantime a Management Board (MB) of the WLCG was scheduled for September 7th , to which the chair of the CRSG was invited. Subsequently, the experiments were asked to submit their reports by this later date. Only one of them did so, with the last usage report handed to the CRSG as late as September 24th, giving the CRSG a short time to analyze and produce a comprehensive summary. Once more we have to kindly ask the experimental collaborations to adhere to the stated deadlines.

The short time systematically available for this committee's deliberations makes our task more difficult and is the ultimate justification as to why this report is submitted to the C-RRB with such short notice, for which we apologize once more.

Following the reception of the reports, referees were assigned to the different experiments and a number of interactions took place. The experiments were provided ahead of the meetings with a set of questions in order to expedite the discussion. As agreed with the ATLAS and CMS management in 2009 the scrutiny procedure for these two experiments is done by a common team of referees, using analogous techniques and methods. This procedure would ensure that a coherent set of principles is applied.

Generally speaking the interactions with the experiments are quite fluid and we thank the respective managements for their openness and collaboration. Thanks are due in particular to Ian Fisk who compiled and summarized the Tier 1 and Tier 2 usage for 2010 on occasion of the MB meeting of September 7th.

A few weeks before the September 7th meeting the CRSG chair held an informal meeting with ATLAS and CMS representatives where some standing issues were raised. These points will be commented on in the section "CRSG work program for 2011". It was not considered necessary to have such informal meetings with ALICE and LHCb as the issues under discussion concern the latter experiments to a lesser extent.

In May 2010 ALICE requested for a clarification of our "Recommendations" section of the C-RRB April 2010 report (CERN-RRB-2010-512). This is commented in the next section.

Interactions with the LHCC

In carrying out the scrutiny of the experiments' requests the scope of this group is largely limited to the implementation of the respective computing models which are periodically reviewed by the LHCC. The evolution of the commissioning of the experiments as well as the implementation of the computing models in successive tests along with a better understanding of their needs have motivated a number of changes, sometimes representing limitations in the original model or assumptions. Obviously there is a gray area where the respective competences of the LHCC and the CRSG overlap and it is not so clear what would represent a change of the computing model or what would just be a natural adaptation of it to the changing circumstances.

The CRSG chair was invited to a LHCC computing mini-review on May 5th. The LHCC has adopted a new shorter format for their meetings that is probably less well suited for the purposes of the CSRG mandate. Following this meeting the CRSG chair wrote to the LHCC Chairman Prof. Terry Wyatt thanking him for the kind invitation extended to the CRSG and raising the convenience or otherwise of the CRSG attending these mini-reviews. In the ensuing mail exchange both chairs expressed their willingness to continue their conversations in order to improve the collaboration between the LHCC and the CRSG on specific issues of computing. However it should be recognized that the roles of the two committees are different.

During the LHCC computing mini-review the ALICE collaboration asked for a clarification of the April 2010 report. This request was transmitted to the CRSG, which had not been approached by ALICE representatives previously. After analyzing the content of the original report it was decided to produce a revised version. According to this, the last two paragraphs on page 20 of document CERN-RRB-2010-052 should be replaced with the following text:

"The same cannot be said for the CPU resources they request. While these fall 10% short at CERN in 2010 (a year for which the scrutiny is already completed) they are projected to be adequate at CERN in 2011. Their sum of T1/T2 CPU requests grows substantially in 2011 and the pledges appear to fall about 30% short before the end of the 2011 CRRB year. This corresponds to the first quarter of 2012 when ALICE will start to process the HI events from the November 2011 run.

We believe the ALICE collaboration will be able to analyze the HI and pp data they will receive in 2010 and the pp data of 2011 given the LCG computing resources currently pledged. The analysis of the HI data to be taken in November 2011 may extend longer throughout 2012 as the current level of pledges is too low for that time period if the current estimates for HI processing hold. We note that the estimates for HI data processing remain largely untested and will not confront real data until the 2010 HI run.

Their request for resources in 2012 will need further study as the collaboration anticipates continued increases will be necessary, at a similar rate to 2011, despite the fact that no new data will be collected by the ALICE experiment in 2012."

We emphasize that the core of the ALICE collaboration computing model lies in Heavy Ion (HI) collisions. The initial HI run will start in a few weeks and therefore a big part of the ALICE computing model is naturally in a more fluid state than that of the other collaborations.

Since the last April 2010 scrutiny no issues appeared for which we thought it was necessary to refer to the LHCC.

On the scrutiny process

The CRSG is now satisfied with the quality and quantity of the information provided by the experimental collaborations. The experiments reports are well documented.

For future reviews we insist that the following good practices be maintained:

- Requirements and respective models should be frozen during the review
- All changes to the models compared to the previous review should be well documented
- All documents should be provided sufficiently early to allow time for the review, a deadline for the revised requirements should be agreed upon well ahead of the final report deadline. For the upcoming April 2011 C-RRB meeting this deadline is **March 1st 2011**.

The implementation of the WLCG MoU should be agile enough to respond to the changing environment we face during the early days of the LHC. In the past a slow reaction has probably had a non-negligible cost for the funding agencies. The CRSG is committed to make recommendations to the C-RRB with the aim of optimizing the resources already invested and adjusting the future ones to the required computing needs.

As the experiments have not produced tentative estimates for 2012 and 2013 we have not been able to provide a first assessment of them yet.

General comments

The comments specific to the scrutiny of the usage in 2010 are deferred to part A of this report. Here we comment on the scrutiny and reviewing process itself.

- The WLCG accounting of Tier 2 resources is improving steadily but it is still insufficient.
- The Installed CPU compared to the pledged capacity in Tier 2 centers is not properly accounted yet.
- A proper accounting of the real CPU capacity in HS06 has been deemed necessary by some collaborations. Some sort of central benchmarking may be advisable in order to ensure the fulfilment of the WLCG pledges.
- Installed disk capacity at the Tier 2 centres is not accounted so far.
- It would be very convenient for the scrutiny that the CRSG carries out periodically to automate the required information through the EGEE portal.
- Greater precision concerning the future LHC schedule (predicted running time, especially for 2013) is needed from the CERN management in order to provide reasonable estimates.
- The experiments are strongly encouraged to adhere to the September 1st and March 1st deadlines to submit their requests or reports to the CSRG.

CRSG workplan for 2011

1.- The CRSG is often supplied by the experiments with simplified spreadsheets that ease the work of the referees. When available these simplified spreadsheets consistently agree with the ones made of our own, constructed from the information contained in the computing models and the written requests submitted to us. However we have repeatedly noted that some experimental requests are systematically higher than the ones resulting from their simplified models or from the CRSG understanding of the respective computing models although typically by no more than 10%. Unfortunately we have never been able to scrutinize in detail the origin of these differences as specific information is not provided. For 2011 we ask the experimental collaborations to improve the modelling and the amount of information so that the estimates of the CRSG can be assumed to be accurate at the 5% level.

2.- It seems natural that some of the assumptions of the computing models should be revised after the first months of data taking. There are several assumptions that have a direct impact on computing requirements and it seems possible, with the acquired experience, to work on an improvement plan to optimize processing times and data sizes and other parameters so as to limit the expected increase in the resource request in coming years. Some of the potentially relevant issues are

- Experiments should make an effort to reduce the raw event size (and the size of all subsequent derived formats) and event processing time by establishing a reduction profile without unduly jeopardizing the physics. The experiments should also be prepared to make some compromises by revising their data distribution policy, reducing the number of copies stored in Tier 1 or Tier 2, for instance.
- Continuous optimization of the MonteCarlo simulation generators and of the data reprocessing times.
- The potential proliferation of different data formats serving the same purposes should be watched closely. Strong overlap between different primary data sets may also be a matter of concern.
- Care should be taken that the worldwide LCG resources are used as much as possible as there may be a tendency by collaborations to place heavier demands on CERN resources or suggest that a larger than originally planned part of their analysis should be done at CERN.
- The CRSG encourages a close collaboration of the different Tier centers with the experiments to achieve an efficient and cost-effective access to data implementing intelligent storage management policies.

We ask the experiments to work on such an optimization and resource-contention plan in the coming months and to submit their plans to us during 2011 so that, after revision by the CRSG, it can be presented to the C-RRB in October 2011. We take note that since the HI runs have not started and analysis of that data will take place during 2011, it would be too premature to ask ALICE for a resource-contention plan.

We acknowledge the receipt of two ATLAS documents that partly deals with request 1.

On the CRSG membership

Bernd Panzer-Steindel from CERN who replaced Jürgen Knobloch, has asked to step down from the CRSG citing conflict of interests. Tony Cass, also from CERN has been proposed as a new member of the CRSG. The appointment of CRSG members is made by the respective funding agencies, according to the terms of the WLCG MoU contingent on the approval of the C-RRB.

Three new members should join the CRSG in order to comply with the mandate of the WLCG MoU and Funding Agencies have been contacted in the pasts weeks. New appointments should become effective after this C-RRB.

PART A

Scrutiny of the WLCG resources utilization in 2010

This report refers, unless otherwise stated, to the calendar year 2010, from January 1st to September 1st . In the April 2011 C-RRB a report for the full calendar year 2010 will be provided. The experiments are kindly asked to report in **March 1st** about the usage made during theprevious calendar year

This report has used the following sources:

1.- Cumulative accounting from January to August for Tier 1s and CERN

http://lcg.web.cern.ch/LCG/accounting/Tier1/2010/august-10/Master_accounting_summaries_August2010.pdf ,

2.- Month-by-month accounting of the CPU delivered by the Tier 2s

http://lcg.web.cern.ch/LCG/accounting/Tier2/2010/

3.- The EGEE accounting portal at CESGA

http://www3.egee.cesga.es/gridsite/accounting/CESGA/tier1_view.html ,

4.- WLCG accounting reports for non-GRID CPU

5.- 2010 pledges as presented to the C-RRB

6.- The Tier-1 and Tier-2 Usage Reports by Ian Fisk, presented to the WLCG MB

http://indico.cern.ch/getFile.py/access?resId=2&materialId=0&contribId=3&sessionId=0&subContId=0&confId=82006

http://indico.cern.ch/getFile.py/access?resId=1&materialId=0&contribId=3&sessionId=0&subContId=0&confId=82006

7.- The documents that the experiments have provided to the CRSG with rather complete usage reports that are appended to this scrutiny.

General comments

The past months of 2010 have witnessed a substantial change in the WLCG usage with respect to 2009 both in qualitative and in quantitative terms. While the 2009 run offered a limited amount of data used primarily for calibration and commissioning processes, 2010 has brought about sizeable amounts of real data valid for physics analysis. As a consequence the experiments have made extensive and intensive use of the WLCG resources.

Generally speaking the experiments' computing models and the WLCG have demonstrated in a remarkably smooth way their capability to record, distribute and analyze the so far moderate, but rapidly increasing, amounts of data delivered to them by the LHC.

The performance of the WLCG throughout the year has been generally regular and without any noticeable difficulties, with periods where usage has been quite intensive corresponding to simulation and reprocessing campaigns and physics analysis activities jumping to high levels just before the summer conferences.

The GRID fabric works reasonably well, data distribution and network performance are excellent, much better that could perhaps have been envisaged some time ago. A similar comment applies to the middleware.

The data placement policy and the detailed computing activities have been quite different from what was envisaged in the computing models (more concerned with a steady state stage) but this is regarded by the CRSG as being reasonable at this moment.

Resources are exceeding the experiments' needs at this point and the experimental collaborations have had substantial headroom that they have employed to increase simulation production, making reprocessing passes more often (this activity is facilitated by the relatively small number of events to reprocess) and making more copies in Tier 1 and Tier 2 to increase accessibility and to ease the access of enthusiastic physicists to the long awaited real data. In fact, Tier 2s large usage is partly due to an increasing demand by individuals or group for non-organized physics analysis, but the overall use is still dominated by Monte Carlo production. By comparison Tier 1s are still underused at this stage but this situation is changing rather quickly.

If the present trend is to persist it may be the case that Tier 2 resources are stretched in a not too distant future. Redistribution of the organized activities (i.e. changing the computing models) may be advisable. Likewise experiments may need to prioritize or organize individual analysis to some extent in the medium term. Data analysis should move from the more complete data formats (RECO, ESD) to derived formats such as AOD and DPD. As in the past proliferation of data formats should be carefully watched. Given the good performance of the networking, and the possibility of staging copies from tape, unnecessary multiplication of copies can be avoided thus saving on disk. For this last commodity, in general the collaborations have been quite conscientious in cleaning/removing orphan and old data.

Both ATLAS and CMS make a commensurate usage of CERN and Tier 1 CPU resources. CERN usage is at approximately 25% of the nominal capacity due to the so far low luminosity and turnaround LHC output. ATLAS is using a larger fraction of their Tier 1 CPU by carrying out simulation production there (this was not contemplated in their computing model), while CMS has been sticking to their original planning.

ALICE usage of CERN resources compared to external resources is still large but it has decreased with respect to the previous report. On the contrary LHCb relies enormously on CERN resources.

The following table describes the degree of usage of the different resources.

October 2010

Resource	**Site(s)**	**Used/Available**
CPU	CERN	22 %
	T1	50%
	T2	92 %
Disk	CERN	71 %
	T1	92%
	T2	Not available
Tape	CERN	52 %
	T1	49 %

The figures for CPU usage correspond to a time average over the year 2010, obtained from averaging the monthly figures; those referring to disk or tape reflect the amount of the usage relative to the installed capacity at the end of the accounting period (August 2010).The installed capacity is calculated assuming in January the 2009 figures. The 2010 capacity was to be installed by 1 June 2010. A linear interpolation has been assumed in between.

For comparison we reproduce the analogous table presented in the April 2010 C-RRB.

April 2010

Resource	**Site(s)**	**Used/Available**
CPU	CERN	26%
	T1	69%
	T2	(see text below)
Disk	CERN	71%
	T1	81%
	T2	Not available
Tape	CERN	41%
	T1	64%

We see that the numbers show similar trends, even though the total volume of resources is now substantially larger. The one exception is that the CPU usage at the Tier 1 external to CERN is now down to just 50%. Other trends are natural to be expected, for instance the proportion of used tape at CERN has grown, which is directly related to the larger amount of data. Likewise the increased usage of disk at the Tier 1 was also to be expected.

It is to be noted that the percentage usage of CPU at the Tier 2 is very high. This number refers to the HS06*hours used/pledged rather than to HS06*hours used/installed (for the items where the installed capacities are centrally accounted the difference is irrelevant). In 2009 CPU usage at the Tier 2s had actually amply surpassed the pledge in 2009 The CPU delivered by the Tier 2s to the experimental collaborations in 2009 was in average 168% of the pledged value. The main reason for this was found to be that resources installed surpassed pledges in the months from May to September because the deadline for 2009 was moved to autumn following the delay in LHC commissioning. This effect has now disappeared. The efficiency is also much higher than expected, contributing to a very large usage percentage.

The time profile of the installed disk capacity at the Tier 2s is not centrally reported and it is not available. The experimental collaborations have tools to know how much disk is available at a given time in a Tier 2, but no central statistics are kept. Generally speaking the accounting tools for the Tier 2s are still not satisfactory and clearly more work on this is recommended.

Efficiencies

The computing TDR estimates the efficiency to be 85% for CPU and 70% for disk in the case of organized (group driven) analysis, reducing to 60% in the case of chaotic (user-driven analysis).

For simplicity it is assumed there that organized activities are carried out at CERN+ Tier1's while chaotic analysis is carried out only at Tier2's. (Actually CERN is a combination of T0/T1/T2/T3 varying in relative percentage from experiment to experiment). We note that the current efficiency at CERN + Tier1s is 80%, close to the reference value, with little spread among the different experiments.

In the current implementation of the computing models only a fraction of the Tier 2 is user/chaotic analysis and this surely reflects in the fact that the percentage of utilization in the Tier2 even in this early stage is well above the theoretical 60% lying at the 76% when averaged over the Tier2 (It was 75% in the April 2010 report). This results in the Tier2 systematically yielding CPU well above the pledge.

This large efficiency is present even if user analysis is taking place in increasing numbers. It seems to be present in the CMS efficiency time-profile as this shows a dip in the period previous to the summer conferences (up to 500 users in mid-June), but it is not dramatic one. This phenomenon is much less marked in the ATLAS efficiency time-profile and completely absent in LHCb, which does not foresee user analysis at Tier 2s . ALICE shows a very irregular efficiency profile with periods where the efficiency plummets to a 35% only. However the average value is still high.

It is interesting to note that LHCb performs only organized activities at the Tier 2 (Monte Carlo production) and the efficiency amply surpasses the reference value of 85%. ATLAS is also at this 85% level on average and reports very active individual and group usage for analysis towards the end of the reporting period. However, organized activities have represented for ATLAS still 80% of the Tier 2 output in 2010.

We ask the experimental collaborations to work in improving this efficiency as much as possible as this represents a potential for optimization of the resources.

In view of these figures we recommend a continuous monitoring and if the tendency to large efficiencies is confirmed we would recommend a revision of the official figure of 60%.

Efficiency of the utilization of the CPU at Tier 2s per experiment in 2010 (left column) compared to 2009 (right column)

ALICE	74%	65%
ATLAS	84%	85%
CMS	62%	66%
LHCb	92%	90%

Disk usage

While the interpretation of CPU usage is straightforward, disk usage is more subtle to analyse. A metric based exclusively on disk occupancy does not account for how frequently the data stored on disk are used. It might well be that disks can be full of data which are rarely accessed and therefore can be stored on tape. We feel that a metric that also takes into account a disk access pattern is more appropriate and also identifies “hot spots", namely data that are frequently accessed. We are aware of the technical difficulties involved and of the timescale needed to define and implement such a metric. In the April 2010 C-RRB we required the experimental collaborations to provide the disk utilization in terms of the various data types involved (e.g. RAW, RECO, AOD, derived data, group data, user data and so on) and how frequently they were changed/replaced on disk in order to be able to assess the efficiency of disk usage. ATLAS has provided a detailed document along the above lines (not included in this report), while the other experiments provide substantially less information. We encourage all experiments to follow suit and provide similar studies. A good understanding of the different data volumes and how often they are accessed will allow them to manage disk more efficiently, distribute copies according to the real needs and eventually save resources..

From the available information so far it is claimed that the 70% efficiency assumed in the computing models is a realistic estimate.

Sharing of the WLCG resources

The following tables give an idea of the degree of utilization by the different experiments of the disk and CPU made available to them through the WLCG. The percentages refer to the fraction of the total mass storage, disk and CPU used per experiment (therefore all columns add up to 100% up to rounding errors). On the first (CERN+Tier 1) table the last column indicates which fraction of the total CPU that a given collaboration has used has been at CERN rather than using the T1's (and, consequently, does not add up to 100%). For comparison the 2009 percentages are shown in a separate table

Percentage of use of the resources by experiment in 2010 (CERN+Tier 1s)

Collaboration	% of tape in T1+CERN used at end of period	% of disk in T1+CERN used at end of period	% of CPU in T1+CERN used	% of which at CERN
ALICE	5 %	6 %	16 %	26 %
ATLAS	35 %	57 %	59 %	10 %
CMS	55 %	29 %	17 %	22 %
LHCb	5 %	9 %	9 %	42%

Percentage of use of the resources by experiment in 2009 (CERN+Tier 1s)

Collaboration	% of tape in T1+CERN used at end of period	% of disk in T1+CERN used at end of period	% of CPU in T1+CERN used	% of which at CERN
ALICE	6%	4%	10%	29%
ATLAS	30%	57%	55%	9%
CMS	61%	34%	26%	35%
LHCb	3%	6%	9%	20%

Percentages of usage have stayed relatively constant. The differences between the different placing and analysis strategies when it comes to practice are manifest and likely to consolidate. Note the large usage of CPU and disk by ATLAS in 2010 (59% and 57%, respectively) of the CERN+Tier 1 resources. On the contrary ATLAS has been the one that, in proportion, has used less the CERN resources (a mere 10%). On the other extreme LHCb has done 42% of their computing at CERN, while CMS has lowered substantially its CERN share. LHCb's large fractional usage at CERN should be monitored.

The following figure contains analogous information, but referred to the CPU usage in the Tier 2s. We see that ATLAS accounts for nearly half of the CPU time used.

Percentage of use of the resources by experiment in 2010 (Tier 2s)

Collaboration	% of disk in T2 used at end of period	% of CPU in T2 used
ALICE	N/A	7 %
ATLAS	N/A	59 %
CMS	N/A	29 %
LHCb	N/A	4 %

A detailed account of the Tier 2 usage is attached to this report. It has been prepared for the CRSG by Ian Fisk whom we thank for his help.

Delivered versus pledged

The overall level of fulfilment of the pledges can be learned from the following table. The figures refer in all cases to the end of the reporting period.

Resource	Site(s)	Available / pledged
CPU	CERN	104 %
	T1	100 %
	T2	At least 92 %
Disk	CERN	100 %
	T1	97 %
	T2	N/A
Tape	CERN	100 %
	T1	100 %

These percentages of fulfilment are of course quite satisfactory. However there are local anomalies that are sometimes compensated by other centers' over-commitment. For instance:

- CNAF delivered only 73% of the pledged CPU
- CNAF and ASGC delivered less than 75% of the pledged disk
- ASGC delivered 71% of the pledged tape.

The large turnout in CPU at the Tier 2 indicated that the percentage installed is actually superior to 100%, however automated accounting of this is still not in place.

The CRSG wishes to state that the recommendations contained in our scrutinies are to the best of our knowledge rigorous. They correspond to the real needs of the experiments for a given LHC live time in the present stage of their computing model implementation. There is no contingency for late delivery or failure to meet the pledges included in our estimates or for less than 100% availability of these resources.

.

Usage by the individual experimental collaborations

Below we provide some tables, ordered alphabetically by experimental collaboration, that give a rather accurate picture of the experiments usage of the resources. Every table is followed by some comments. The data has been provided by the collaborations themselves and cross-checked whenever possible.

ALICE

Resource	**Site(s)**	**Pledge [1]**	**Alice usage [2]**	**Alice usage CRSG [3,4,5]**
CPU/kHS06	T0+CAF	46.8	15 (30)	17.9 (37.8)
	T1	45.6	23.4 (62.6)	30.8 (73.3)
	T2	52.6	21.0 (54.4)	31.5 (64.9)
Disk/TB	T0+CAF	5500	610 + 520 buffer	781
	T1	6122	986 + buffer	1317
	T2	4326	1090	?
Tape/TB	T0+CAF	6300	1551	1754
	T1	8485	620	418

[1] Pledges from April 2010 RRB.

[2] From Alice "Computing Resources Usage in 2010", 28 Sep 2010, covering Jan-Aug 2010. CPU values in parentheses are averages over July and August.

[3] CERN and Tier 1 data are from the WLCG Tier 1 Accounting Summary for August 2010 covering Jan to Aug 2010. Tier 2 CPU is from the Tier 2 spreadsheet supplied to the CRSG by Ian Fisk.

[4] CPU is total normalised wall time in kHS06 days from Jan-Aug, divided by the number of days, 243. Values in parentheses are usage in Aug divided by 31 days.

[5] Disk and tape values are at the end of the period. We were not able to make our own estimate of T2 disk use.

Comments on the ALICE usage report

The experiment aims to collect 1.5e9 pp events by the end of the pp run, as stated in their resources request (this compares to 1e9 pp events in a "standard year" of the computing model). The pp event size has been growing as the LHC luminosity increases, but is still below the size in the computing model: 0.9 MB per event in LHC10e running compared to 1.1 MB in the model. About 1 PB of raw plus derived data has been accumulated so far in 2010.

Monte Carlo generation includes heavy-ion events in anticipation of the PbPb run to come. For pp MC the event size is about 60% larger than in the computing model. However for AA events the collaboration is investigating storing a much-reduced format without the raw data, nearly an order-of-magnitude smaller at 35 MB compared to 306 MB in the computing model. Some raw AA MC data will be kept, but the reduction in the storage requirement should propagate into future resource requests.

CPU capacity measured from wall-time was underused on average, but has exceeded the pledges in the last two months of the period at T1 and T2. Usage at CERN (T0 plus CAF) remains low but is growing rapidly now; we recall that T0 capacity in the computing model is dominated by the heavy-ion reconstruction requirement which is yet to be tested. CPU efficiency measured by the ratio of normalised CPU hours to normalised wall hours was low in May and very low in June 2010 and has been consistently low at CERN (driven by low values for “nongrid” usage). ALICE noted to us that the dip in May and June arose because no Monte Carlo was ready to run then. CPU efficiencies at T1 and T2 for Jan-Aug are just over 70%, which is consistent with the mean of the WLCG standard efficiencies for scheduled and chaotic analysis.

Disk and tape storage usage are below the pledges, especially for T1 tape, but this has been accumulating only LHC data (no cosmics). The AA run at end of year should increase usage substantially since in a “standard year” of the computing model the raw AA data from one month is anticipated to be more than the raw pp data from the remainder of the year.

ALICE resource usage is generally low compared to the pledges but rising rapidly now. For a full picture we still await the first test of the computing model against real heavy-ion data.

ATLAS

Resource	Site(s)	Required [1]	Pledged [1]	Used [2]	Used/ Pledged	Average CPU efficiency
CPU (kHS06)	T0+CAF	67	67	24 (31)	36% (46%)	47% (55%)
	T1	192	211	117 (106)	55% (50%)	83% (80%)
	T2	240	215	174 (167)	81% (78%)	84% (79%)
Disk (TB)	T0+CAF	3900	3900	3145	81%	-
	T1	21900	22018	18007	82%	-
	T2	20900	21238	12000	57%	-
Tape (TB)	T0+CAF	8900	8900	6426	72%	-
	T1	14200	15372	7814	51%	-

The figures for CPU usage correspond to a time average over the first 8 months of 2010, those referring to disk or tape reflect the amount of the resource installed at the end of August 2010.

[1] required and pledged resources from April 2010 RRB (http://lcg.web.cern.ch/lcg/Resources/WLCGResources-2009-2010_12APR10.pdf)

[2] storage information from WLCG accounting summaries (http://lcg.web.cern.ch/lcg/accounting/Tier1/2010/august-10/Master_accounting_summaries_August2010.pdf), CPU information from spreadsheets supplied to the CRSG by Ian Fisk

Comments on the ATLAS usage report

ATLAS has accumulated 2.3 million live seconds of running until August 31st, 2010. The performance of the ATLAS computing system was generally smooth throughout the whole period; more intensive use is observed, corresponding to simulation production and reprocessing campaigns and to user activities just before the summer conferences.

The resources made available to ATLAS were actually larger than they requested in September 2009 due to the shift in the commissioning of the LHC. In general, resources are exceeding the experiment's needs at this point, resulting in a large overhead which was successfully exploited in order to increase the number of simulated events, of reprocessing passes, and of data copies across the Tier1 and Tier2 ensembles with respect to the nominal ATLAS computing model. Data placement and computing activities were quite different from what was foreseen in the computing model, but this is reasonable at this stage of data taking.

CPUs at CERN were underused on average (about 35% of the ATLAS nominal capacity), but saturated when LHC was giving physics collisions. CPU usage at Tier 1 sites was significantly higher, but still at the 55% level, with sites exceeding their share in some periods and usage dominated by simulation production and physics groups processing ESD events. Tier2 were used also at about 100% of their CPU capacities, with usage also dominated by simulation production but with physics groups and users ramping up to about 35% of the total usage in the last 3 months. By profiting from the resources made available to them, ATLAS could produce 60% more simulated events than previously foreseen and did not need to generate fast simulations.

CPU efficiency has been generally high at Tier1s, and close to the 85% efficiency assumed by WLCG. At CERN, the average is about 40% but with there has been a steep increase in the last three months and is now similar to the Tier1 CPU efficiency.

ATLAS distributed data in 2009 and 2010 with more copies (termed 'secondary') than described in the computing model in order to facilitate physics and performance studies, again profiting from the extra capacities available. These secondary datasets (mostly ESDs) are being deleted when space is needed. ATLAS is making progress towards a dynamic model, where data access patterns are monitored, less popular data are automatically removed and dynamic replicas of data are triggered on Tier2 sites whenever needed. We welcome this development as it goes in the direction of making their computing model more sustainable at nominal conditions.

We encourage ATLAS to continue moving data analysis from more complete data formats (ESD) to derived formats (AOD, DPD). As in the past, proliferation of data formats should be carefully watched. We note progress in reducing the average CPU time requested to simulate an event and encourage progressing further on that.

Resource	**Site(s)**	**Requested [1]**	**Pledged [1]**	**Used [2]**	**Used/ Pledged**	**Average CPU efficiency**
CPU (kHS06)	T0+CAF	96.6	96.6	16.6	17% (19%)	37% (35%)
	T1	100.5	103.5	33.5 (36.7)	32% (35%)	71% (70%)
	T2	195	196	149	76% (80%)	62% (57%)
Disk (TB)	T0+CAF	4100	4100	2456	60%	-
	T1	13400	12183	8393	69%	-
	T2	9200	13627	7500	55%	-
Tape (TB)	T0+CAF	14600	14600	7133	49%	-
	T1	23300	23677	15439	65%	-

The figures for CPU usage correspond to a time average over the first 8 months of 2010, those referring to disk or tape reflect the amount of the resource installed at the end of August 2010.

[1] pledged resources from April 2010 RRB (http://lcg.web.cern.ch/lcg/Resources/WLCGResources-2009-2010_12APR10.pdf)

[2] storage information from WLCG accounting summaries (http://lcg.web.cern.ch/lcg/accounting/Tier1/2010/august-10/Master_accounting_summaries_August2010.pdf), CPU information from spreadsheets supplied to the CRSG by Ian Fisk

Comments on the CMS usage report

The performance of the CMS computing system was generally smooth throughout the whole period.

Also for CMS, the low live time of LHC has manifested itself as a lower utilization of the Tier-0 processing resources and faster turn around for data reprocessing at the Tier-1s. CPU usage was 17% CERN, about twice at Tier1s.

Ten partial or complete processing passes over the data were performed in the first 6 months of running. The resulting data volume could be accommodated since the size for raw and reco data has been much smaller than planned for. We note that the AOD size grew substantially with respect to what initially foreseen and we recommend keeping it under control in the future. The overlap in primary datasets due to trigger (40%) should also be monitored and reduced in 2011 to more manageable levels.

The CPU utilization of the Tier-2s is at the 75% level. Tier-2s have been heavily used for simulation and analysis. The heavy use of the Tier-2s is an indication of the success of moving people to use distributed computing resources.

CPU efficiency is slightly lower than the expected values at Tier1s and Tier2, but improving.

CMS has accumulated about 700 million collision events in 2010, and about the same number of simulated events, corresponding to a data volume of 2PB. Data placement of this as well as previous data and simulations at Tier1 and Tier2 reflect in disk occupancies of 95% and 55% of the available disk space, respectively.

Disk accounting from CMS is not particularly detailed at this point. We understand that analysis was facilitated by distributing more data copies in computing centers than previously anticipated, benefiting from the overhead due to lower running time. We encourage CMS to implement a detailed monitoring of disk access in order to understand usage patterns and possibly establish a data distribution model to optimize resource usage.

LHCb

Site	**CPU (kHS06)**	**CPU (kHS06) (last 3 months)**	**Disk (TB) Request**	**Disk (TB) Used**	**Tape (TB) Request**	**Tape (TB) Used**
CERN	21	3.2 (4.9)	1220	850	1500	788
Tier-1	41	10.3 (18.2)	2870	2064	2800	835
Tier-2	36	12.7 (19.7)	20	1	0	0

Comparison between the 2010 requests (labelled “April 2010” in the requests approved by the RRB) and the present consumption (9 months). The CPU power over the last 3 months is indicated in brackets and demonstrates the significant increase of CPU needs during the data taking.

* 70 Tb front-end (buffer) disk space is subtracted from each T1 disk request in order to compare the usage to the effective storage area.

Comments to the LHCb usage report

The usage of the computing resources by LHCb displays a rather healthy situation, The experiment is successful in using the resources and adapting to the unexpected pile up. It should be noted that the recent physics output of the LHCb collaboration is also an indication that the computing is in good shape and is able so far to process and analyse efficiently the collected data.

The usage of resources in 2010 as of Oct. 4th is presented in the table. The detector is in good shape and the data collection efficiency is now close to 92%. The RAW data distribution from T0 to T1s proceeds according to the plan: 116 Tb of data were shared among the 6 T1s outside CERN.

The running conditions at LHC with high intensity per bunch lead to an average collisions per bunch up to 1.5-2 while the nominal conditions assume 0.4. As a result, significant increases in the event size and in the CPU needs to reprocess each events were observed. The event size increased from 80 kB to 140kB. The reconstruction algorithms were tuned to cope with the high occupancy. The CPU time scales linearly with the average number of collisions per bunch crossing μ, doubling when μ increases from 0.4 (nominal) to 2 as reached during 2010. In addition, looser cuts during the stripping were applied as expected for the first studies using real data.

These effects (event size, CPU, loose stripping) lead to high disk occupancy and to a need to decrease the number of copies in T1s in order to save resources. The average replication factor was in fact around 6 instead of 7. The replication will be further reduced to 3 and some older reprocessing (Reco4) will be removed in order to free resources for the incoming data. The stripping rate is continuously reduced to achieve the design 10% and strategies to use the μDST format for some streams are under study.

The Monte Carlo production followed the plan but observed also an increase in activity. The data distribution in 3 replicas (one of which in DST version resides on disk in a T1) proceeded according to the plan with an average replication rate of 2.6.

A campaign of processing of data and Monte Carlo production with best parameter sets deduced from data is expected to start by end 2010 to be ready for spring 2011 conferences.

The CPU resources usage is not continuous and the installed power is used on (frequent) processing bursts. The data flow is obviously dependent on machine delivery. This corresponds to a variety of Monte Carlo production which are expected to be available in a short time. The data is used to tune the simulation and this process logically lead to successive processings of Monte Carlo samples In addition, a modulation of the analysis flux is also naturally expected. As a result, the difference between the average CPU usage and the peak usage is a factor of 4 in T1s. It is likely that rearrangements will occur as the experiment progress through data taking, but the CPU peak power is expected to continue be an important parameter for LHCb.

The usage of CPU is distributed according to pledges for the regular productions (reconstruction, stripping) and reaches about 45% for CERN. The analysis jobs are in turn distributed according to the site availability and reaches about 45% at CERN as well. According to LHCb experts, this is a sign of the availability of the CERN resources when compared with other T1s.

The LHCb computing model is robust and was able to incorporate parameter adjustments beyond the anticipated variations. The usage of the resources is satisfactory, given the running conditions. The used CPU power is different in average from the pledges but incorporates the necessary contingency for burst of data and MC production. As already noted by CRSG, LHCb would benefit from a mutualisation of resources at CERN.The usage of tapes is significantly lower than expected mainly due to less running time in 2010. The disk occupancy is relatively high and is a matter of concern. LHCb experts took correct and timely actions (clean-up older sets, reduction in replication, stripping optimization) to cope with a possible limitation are are confident that the data taking, the MC production and the reprocessing foreseen by the end of the year will be covered within the available resources. Since the data collection rate is continuously increasing, a better view will be available by end 2010 for the needs for 2011 and beyond. For instance, it is likely that the disk space in the T1s and the corresponding data distribution strategy will suffer modifications.

PART B

Experiments usage reports and Tier 2 report

ALICE
Yves Schutz
Tuesday, September 7, 2010

COMPUTING RESOURCES USAGE IN 2010

The usage of computing resources by ALICE in 2010, from January to September, is discussed. An update of the computing model is reported. Detailed data on the usage of CPU cycles and disk storage are extracted from the WLCG monthly accounting reports for Tier 0 and Tier 1s[1] and for Tier 2s[2] and from the EGEE accounting portal[3]. The ALICE accounting report, including the non-LCG sites, can be retrieved from the ALICE MonALISA web pages[4].

Data collection

The profile of data recording during the period January 2010 to September 2010 is displayed in Fig. 1. This period corresponds to four LHC running periods, defined as the time between two LHC technical stops: LHC10b, LHC10c, LHC10d and LHC10e.

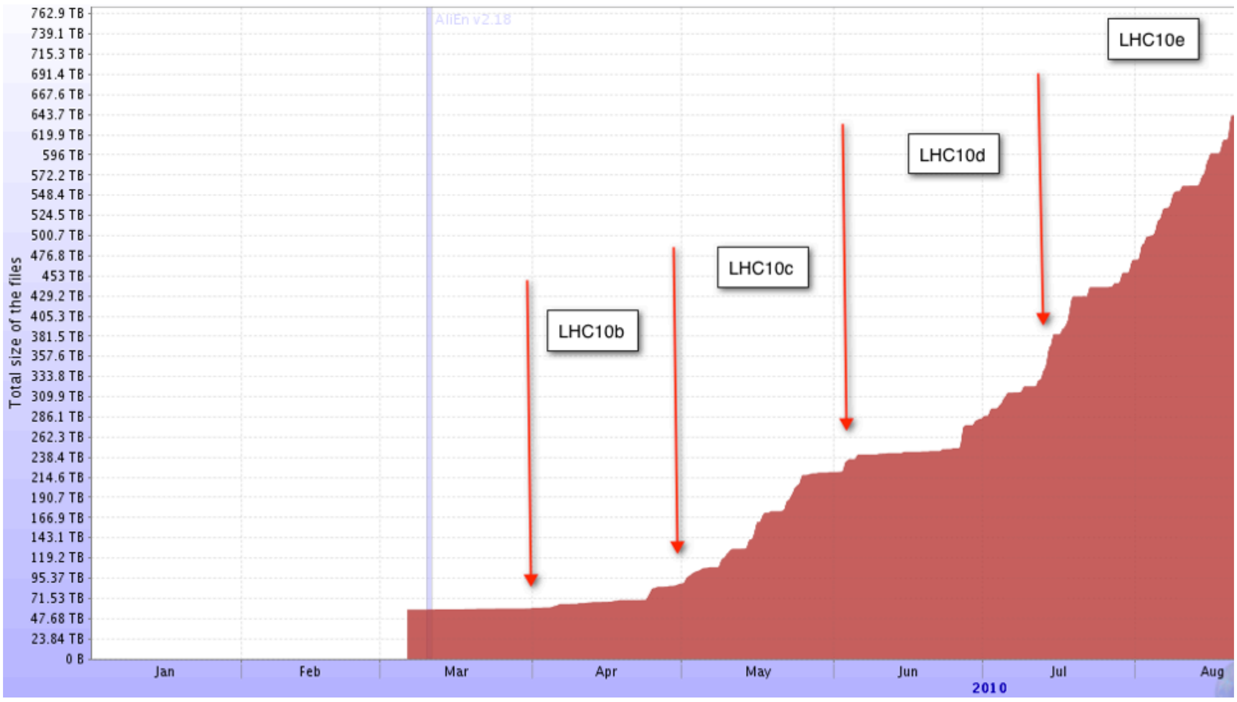


Fig. 1: *ALICE data taking profile in 2010 until September. The three LHC running period are marked.*

---

[1] http://lcg.web.cern.ch/LCG/accounting/Tier1/2010/june-10/Master_accounting_summaries_June2010.pdf

[2] http://lcg.web.cern.ch/LCG/accounting/Tier-2/2010/july-10/Tier2_Accounting_Report_July2010.pdf

[3] http://www3.egee.cesga.es/gridsite/accounting/CESGA/egee_view.php

[4] http://pcalimonitor.cern.ch/reports/

The number of events recorded by ALICE during these periods is reported in Tab. 1. The average data taking efficiency (used luminosity over delivered luminosity) since the beginning of the *pp* run is of the order of 80%. Until September, 694×10^6 minimum bias triggers and 63×10^6 rare triggers (muon and high multiplicity triggers) have been recorded (not including tests, calibration runs etc...). The objective is to collect 1.0×10^9 minimum bias events (69% today) plus 0.5×10^9 (13% today) rare triggers by the end of the 2010 *pp* run.

Tab. 1: *Number of events, Minimum Bias (MB) triggers with most of detectors (in parenthesis all MB events) and rare triggers, reconstructed by ALICE in 2010 until September during three LHC periods and two beams settings.*

		Number of events ($\times 10^6$)				
		LHC10b	LHC10c	LHC10d	LHC10e	Total
pp @ 0.9 TeV	MB	-	8.5	-	-	8.5
	Rare	-	0.024	-	-	0.024
pp @ 7.0 TeV	MB	75	391	232	268	694
	Rare	0.5	4.2	12.8	45.4	62.9

During the LHC technical stops, ALICE has recorded cosmic and calibration data.

The *pp* event size was on average 0.3 MB for period LHC10b-c (to be compared to 0.2 MB per event without pileup assumed in the computing model) and rose to 0.90 MB for LHC10e where the higher luminosity generates pileup (to be compared to 1.1 MB per event with pileup assumed in the computing model).

The total amounts of 2010 raw data stored in the CASTOR mass storage system at CERN and replicated once in Tier 1's mass storage are reported in Tab. 2.

Tab. 2: *Amount of raw, reconstructed and analysis data stored by ALICE in 2010 until September during four LHC periods. The amount of reconstructed data for LHC10b-c correspond to 2 passes, and to 1 pass for LHC10d-e and includes a factor 3 to take into account replica.*

Data type	Amount				
	LHC10b	LHC10c	LHC10d	LHC10e	Total
Raw (TB)	26	111	186	300	623
ESD (TB)	42	180	57	90	369
AOD (TB)	1.8	6	5.6	18	31.4
Total (TB)	69.8	297	248.6	408	1023.4

Data processing

To increase the stability of the code over a running period, we have changed the AliRoot release policy. As a consequence we have adapted the processing strategy to the *pp* running condition. This change has no impact on the overall resources usage.

Every week a new tag of the AliRoot version is published exclusively correcting bugs. A new version of the code including new features is published on a monthly basis. This has lead us to the following processing strategy:

- Pass 1 reconstruction is performed in the Tier0 exclusively, and is started as soon as data are registered in CASTOR and AliEn. The code is updated and a new tag is published weekly.
- Pass 2 reconstruction is started in the Tier1s about 2 weeks after the end of a period, the code and the condition data are frozen for a given pass. Data from two consecutive periods can be reconstructed with different version of the code.
- Pass 3 reconstruction is started in the Tier1s one month after the end of the LHC run, the code and the condition data being frozen. This last pass provides reconstructed data with a unique version of AliRoot and a unique set of condition data.
- An analysis train, the QA train, is started for each ALICE run in the same job as the reconstruction. This provides an immediate feedback on the data and reconstruction quality and the AODs (filtered ESDs) over which the PWG trains and the end user analyses are run.

All tasks are performed on the Grid. Pass 1 reconstruction runs exclusively in Tier0, additional reconstruction passes run exclusively in Tier1s. PWG trains, end user analysis and Monte-Carlo productions run indifferently in Tier0, Tier1s and Tier2s.

The efficiency (defined as the fraction of events successfully reconstructed, the Grid efficiency being folded out) of pass 1 reconstruction is larger than 98% and the efficiency for further passes larger than 99%.

CPU power

The average power required to reconstruct one event is 50 HEP-SPEC06×s compared to the value of 45 HEP-SPEC06×s used in the computing model (including the 85% CPU efficiency factor).

ESDs and AODs size

The size of the ESD has been optimized since the beginning of the year. The debugging information needed in the initial phase of running have been progressively reduced to reach an ESD size equal to 10% of the size of raw data as foreseen in the computing model. ESDs are replicated three times in various Storage Elements in Tier1s and Tier2s. The total amount of ESDs data produced, including the replication factor, are reported in Tab. 1. The size of the raw AODs (filtered

ESDs) is 1% of the size of the Raw data which is the value in the computing model. They are replicated six times in Tier1 and Tier2 storage elements to optimize the data analysis on the Grid.

The profile of raw data processing is displayed in Fig. 2 in terms of number of concurrently running job. The average amount of concurrent analysis jobs (PWG trains and end users analysis) is since the beginning of the year equal 1150 jobs with maxima of 3800 jobs for about 250 regular users. Since the beginning of the year 170 trains have run over 2009 and 2010 real and Monte-Carlo data.

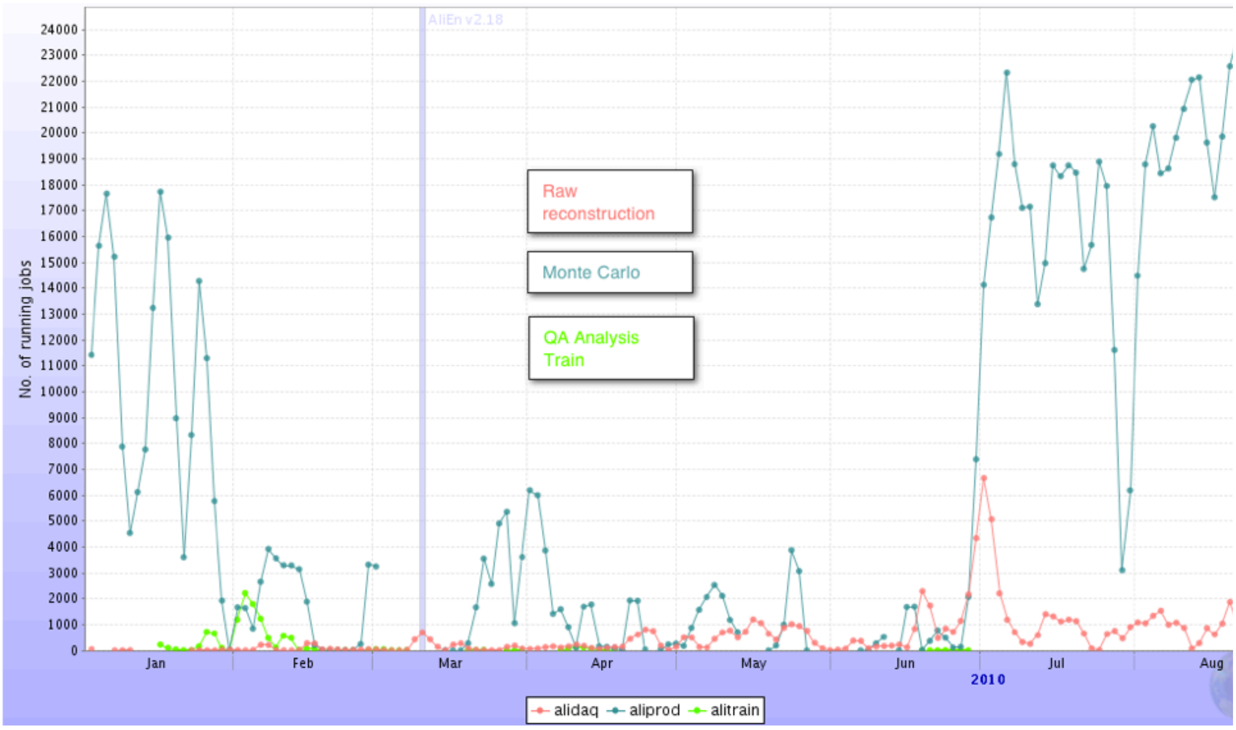


Fig. 2: *Profile of running jobs in 2010 for the reconstruction of raw data, the Monte-Carlo production and the QA analysis train.*

Monte-Carlo production

For ***pp*** runs, one Monte-Carlo event is generated and reconstructed for every reconstructed real event as foreseen in the computing model. These productions represent the major contribution to the overall Monte-Carlo production. Additional productions were performed to study dependencies with the event generators, to study the performance of various transport models, for rare signals and for specific detector needs. Heavy-ion data were generated as well in anticipation of the PbPb run scheduled at the end of this year.

CPU power

The average power needed to reconstruct one event is 620 HEP-SPEC06×s for pp collisions at 900 GeV, 840 HEP-SPEC06×s for pp collisions at 7 TeV, compared to the value of 340 HEP-SPEC06×s used in the computing model (including the 85% CPU efficiency factor). We are working on optimizing the code to reduce this CPU power. For AA Monte-Carlo, we need 123K HEP-SPEC06×s compared to 127K HEP-SPEC06×s used in the computing model.

Data size

The overall size of the data generated by the Monte-Carlo is 0.67 MB per pp event and 35 MB per AA event. These numbers are 0.41 MB and 306 MB in the computing model. In the computing model we had foreseen to store in addition the raw data format. This is presently done only for a few special productions.

Accounting

The WLCG resources used by ALICE in 2010 are summarized in Tab. 3. Data have been collected from the WLCG accounting reports, the EGEE accounting portal and the MonALISA accounting. The average values over 8 months since January are reported together with the values averaged over the past two months (July-August).

These numbers can be compared to the pledged CPU resources[5], 47K HEP-SPEC06 by Tier0, 46K HEP-SPEC06 by Tier1s, and 53K HEP-SPEC06 by Tier2s. In the past months we have used all the resources available to ALICE in Tier0, Tier1s and WLCG Tier2s.

Concerning storage, so far we have used at the Tier0 15% of the installed disk space and 22% (cumulative) of the pledged tapes. In the external Tier1 sites, we have used 45% of the installed disk space and 8% (cumulative) of the pledged tapes. In the Tier0 and Tier1s disk usage we have not counted the disk used as buffer of the mass storage system. In the WLCG Tier2s we have used so far 49% of the installed disk space. The difference between the mass storage used in Tier0 and in the Tier1s is explained by the facts that one one hand data replication in the Tier1s has only started with the start of LHC (cosmics data have not been replicated) and on the other hand only interaction data are replicated, excluding technical runs, calibration data and detector tests.

In addition to the T2 resources accounted by WLCG, about 20% of the total used CPU resources and the total used storage is provided by non WLCG sites.

As a reminder, the storage needed for pp represents about 30% of the total storage needed in 2010 including the pp and the AA run.

Tab 3: *Usage of WLCG resources in 2010. CPU and Disk are average over 8 months (in parenthesis the average over July and August), tape is the amount used by September 2010. T0 and T1 and T2 data are taken from the WLCG and MoALISA accounting.*

	TO+CAF	**T1**	**T2**
CPU (K HEP-SPEC06)	15 (30)	23.4 (62.6)	21.0 (54.4)

[5] http://lcg.web.cern.ch/LCG/Resources/WLCGResources-2009-2010_12APR10.pdf

	TO+CAF	T1	T2
Disk (TB)	610+520 (tape buffer)	986+tape buffer	1090
Tape (TB)	1551	620	0

The fraction of CPU power used for raw data reconstruction, data analysis (analysis train and end users analysis) and Monte-Carlo productions is 6%, 13% and 81% respectively. In the computing model these numbers are 15%, 30% and 55%.

COMPUTING RESOURCES ESTIMATE FOR 2012-2013

The required resources for 2012 and 2013 needed to process the data collected in 2010 and 2011 are summarized in Tab. 4. These numbers have already been discussed with the CRSG referees in 2009 and have not been changed them.

In 2013, the ALICE data taking scenario will be similar to the scenario of 2010: acquisition of 1.0 $\times 10^9$ minimum bias events plus 0.5 $\times 10^9$ rare events in the ***pp*** run and 1.0 $\times 10^8$ events in the ***PbPb*** run. Detailed requirements for 2013 will be calculated after the 4 experiments have adopted a common running scenario of the LHC.

Tab. 4: *Resources required for RRB years 2011 and 2012.*

RRB year	Start	End	CPU (KHEP06)				DISK (TB)				Tape (TB)	
			T0	CAF	T1	T2	T0	T1	T2	CAF	T0	T1
2009		05/10	7.6E+00	5.4E-01	4.5E+01	7.0E+01	2.3E+03	1.3E+03	1.8E+03	0.0E+00	2.0E+03	2.6E+03
2010	06/10	03/11	4.0E+01	1.4E+01	4.5E+01	7.0E+01	3.3E+03	5.2E+03	4.9E+03	1.6E+02	4.6E+03	7.5E+03
2011	04/11	03/12	4.8E+01	1.4E+01	1.2E+02	1.6E+02	5.8E+03	8.5E+03	6.6E+03	5.4E+02	6.8E+03	1.3E+04
2012	04/12	03/13	5.1E+01	1.4E+01	1.5E+02	1.6E+02	7.1E+03	1.0E+04	8.2E+03	9.1E+02	7.5E+03	1.7E+04

Report on ATLAS Computing Resource Usage in 2010

This document shows present actual ATLAS usage of computing resources compared to the predicted needs. The predicted needs and the format of the tables below are from the document "ATLAS Computing and Resource Needs for Data-taking in 2010-2012" dated 17 March 2010 as submitted to the RRB.

While reflecting actual usage patterns, the figures presented are for usage during initial data taking, and in many ways do not reflect the usage patterns expected for steady high-luminosity running in years to come. Some important differences in actual conditions to date are: machine live time is still low; data has been collected (and simulated) with a variety of LHC and experiment conditions; analysis activities have focused more on detector commissioning and performance studies rather than physics analysis of large data samples; and simulation samples have focused on minimum bias physics and standard model processes rather than more complex beyond-the-standard-model searches. Consequently, in analyzing these figures we are able to gain many insights into our computing model; however, we are not yet at the stage of experience that we are able to draw conclusions about how the computing model might evolve and mature.

Scale Factors: We have accumulated 2.3 million live seconds of running so far (31 Aug.) in 2010. We based our resource request on 7.3 million live seconds this year. By live-second accounting we are 0.32 of the way through the year. Note, of course, that the luminosity delivered by the LHC is not constant, it is increasing rapidly. By calendar, we expected to run from March through October with pp collisions so, we are 3/4 of the way through the year. These scale factors are used in the tables below.

In the usage tables below, the **New 2010** column is copied from the resource request document cited above, the **Present** column shows the usage up to 31 August 2010. For disk usage, columns labelled **Present (Primary)** show the amount of space in the Present column that is primary in the ATLAS Distributed Data Management (DDM). Data marked primary are those datasets that are distributed copies of the data that were originally foreseen in the Computing Model. Secondary data are additional copies of the data on disk in any of the Tier-1 or Tier-2s put there to facilitate analysis. The secondary data is deleted whenever space is needed.

Glossary:
Primary - distributed copies of data that were originally foreseen in the computing model.
Secondary - extra distributed copies of data that are beyond the computing model, placed at sites where it facilitates physics and performance studies.
ADC - ATLAS Distributed Computing.
n/a – Not applicable.

Table 1. *Event sizes, samples, processing times and other input parameters for resource calculations.*

LHC and data taking parameters		New 2010	Present	Comment
Rate [Hz]	Hz	200	350	[1]
Run [days]	Days	330	150	
Time [sec]	MSeconds	7.3	2.3	
Real data	MEvents	1460	816	
Simulation	MEvents	600	1000	[2]
Fast Simulation	MEvents	600	0	[3]
Fast Sim kept	MEvents	300	0	
Event sizes				
Real RAW	MB	1.6	1.6	
Real ESD	MB	0.8	1.5	[4]
Real AOD	MB	0.15	0.15	
Sim RAW	MB	2	3	[5]
Sim ESD	MB	1.1	2	
Sim AOD	MB	0.18	0.20	[6]
CPU times per event				
Full sim	HS06 sec	6000	4100	[7]
Fast sim	HS06 sec	400	325	
Real recon	HS06 sec	80	65	[8]
Sim recon	HS06 sec	135	160	[9]
Group analysis	HS06 sec	20		
User analysis	HS06 sec	0.2		

Comments

1. **Trigger Rate:** Because the time used for data taking was initially less than the performance figures used for the resource calculations, we were able to allow a higher trigger rate corresponding to a 400 - 600 MB/s data rate. In spite of this higher data rate, the total accumulated data volume is still below the volume that was assumed in the calculation for the resource request. The higher trigger rate directly benefitted our physics program and included calibration triggers, triggers for detector performance studies, triggers for trigger studies and triggers with less trigger bias.
2. **Simulation:** We have fully simulated more events than originally planned. The need to simulate more events was driven in part by the need to simulate at 3 energies (0.9, 2.36 and 7 TeV). We were able to simulate more events than projected because we had more CPU cycles available, largely from opportunistic resources at T2 and because of the lower number of events recorded (thus quicker re-processing). The opportunistic resources at T2 are shown in the WLCG accounting where many of our T2 centres contributed 50% more CPU cycles than pledged. The additional simulation was driven by our physics analysis needs.
3. **Fast Simulation:** There is still ongoing validation of our fast simulation, so no big campaign has started. At present, full simulation is able to meet our requirements.
4. **Real ESD:** The ESD size/event is bigger than our estimate by a factor of two. This is a concern and we are working on reducing the size. The current large size is attributable to two main factors. One is the initial_beam trigger setup (for simulation) causing a large number of trigger objects to be written out, the other one is that for commissioning purposes ESD have been increased with objects (e.g. silicon clusters, calorimeter raw information for cells above a given threshold) allowing re-reconstruction and commissioning studies from ESD (for detector and reconstruction algorithms tunings).

5. **Sim. Raw:** This data is the Monte Carlo hits plus 50% of the produced raw data objects (RDO). The RDO are mainly used for trigger studies.
6. **Sim AOD:** The AOD is slightly larger than our model in the early running because we are doing a lot of trigger and detector performance studies.
7. **Full Simulation time:** This figure is based on our actual simulation running over a sample of 100 million events containing many minimum bias events. Minimum bias events take 3 times less time to simulate than more complicated events like SUSY. Consequently, more representative samples of simulation of standard model and BSM events require more CPU time.
8. **Real recon:** We expect this to grow as the recorded events become busier on average as we tighten the trigger due to increased luminosity, and eventually due to increased pileup.
9. **Sim. Recon**: The increase of MC reconstruction time can be attributed to two factors: (1) The quoted time now includes digitization. (2) Trigger simulation is run as part of reconstruction. It presently takes 1/3rd of total CPU, in part because we have been running with open trigger settings that have increased CPU time.

Table 2. ATLAS Resources Summary Table

CPU [kHS06]	*New 2010*	*Scaled by calendar*	*Scaled by live seconds*	*Present*	*Present Primary/ comment*
CERN	74			63 available	~30% used*
Tier-1	178			190 available	~50% used*
Tier-2	226			more than pledged (240)	~120% used*
Disk [PB]					
CERN	4.7	3.5	1.5	2.4	0.8
Tier-1	22	16	7	18	7.6
Tier-2	24	18	8	12	3
Tape [PB]					
CERN	8.9	7	3	7.3	Custodial
Tier-1	18	13	6	5.6	Custodial

* From WLCG accounting reports.

CPU: Detailed CPU usage numbers can be seen in the WLCG accounting plots generated for all 4 experiments. Usage is determined per month as fraction of used over pledged. For T1s, this includes substantial downtime for the ASGC centre, although it is fully functional now.

The total installed CPU capacity in the Tier-2s is not officially known, but some sites have made available more than pledged, affording more cycles than pledged.

Disk: We do not expect disk usage to scale with live time as we maximally fill disks with useful data as soon as the disks are commissioned. This extra filling is secondary data that is automatically deleted when the space is needed. Most of the data on disk at CERN is secondary and used for calibration, alignment and performance studies.

Tape: The real data on tape scales with live time, but the simulation component does not. Our usage of tape is above the live time scaled requests mostly due to the extra simulation as well as legacy data that we do not erase from tape.

Tier 1

Table 3. ATLAS Tier 1 Disk Usage

Tier-1 Disk [PB]	***New 2010***	***Scaled by calendar***	***Scaled by live seconds***	**Present**	**Present Primary**	**Comment**
Current RAW data	3.9	3	1.2	9.1	3.6	This is the sum of "Current RAW data" and Real (ESD,AOD,DPD) data
Real (ESD, AOD, DPD) data	6.5	5	2			
Simulated data	4.1	3	1.3	7.2	4	
Calibration and alignment inputs	2.2	1.6	0.7	~0	n/a	
Group data	1.4	1	0.4	0.4	n/a	
User data	0.6	0.4	0.2	0.6	n/a	
Cosmic ray data	0.2	0.15	.1	~0	n/a	
Processing and I/O buffers	2.9	2	0.9	0.8	n/a	
Total	***22***	**16.5**	**7**	**18**	**7.6**	

Comments on Tier-1 disk

Current RAW data and **Real (ESD, AOD, DPD) data:** The primary data on tier-1 disk should scale with live time, and our actual usage reflects this behaviour. That is, our primary data on T1 disk (3.6) is consistent with the 3.1 scaled request, considering the temporary operation with an increased trigger rate. The secondary component is extra replicas that facilitate physics and performance studies. At present, the secondary component is mostly ESD, which was needed during the early commissioning phase. We expect this usage pattern to evolve, and we have a flexible plan to fully occupy our disks with the most useful data.

Simulated data: We have simulated more data than originally planned for this year. We simulated events at 3 different energies: 900 GeV, 2.36 TeV and 7 TeV. Because we have recorded less real data than expected to date, more tier-1 CPU and disk capacity has been available to support this additional simulation.

Calibration and alignment inputs: At present, the resources at the CAF at CERN have been sufficient to perform all calibration and alignment tasks. Consequently, Tier-1 disk space for calibration and alignment inputs has not yet been utilized. (Use of the CAF rather than Tier 1 for calibration and alignment has been advantageous during the commissioning period, due to the need for frequent software updates and validation.) We continue to plan for the need for these inputs in the future, when routine calibration and alignment tasks move away from the CAF.

Group data and **User data:** Group production is ramping up gradually. Some of this group activity is currently being done by individuals using their own grid certificate, so it looks like user analysis to our monitoring, and its output is misclassified as user data. The ADC operations group is actively working with physics and performance groups to automate group production using our central production system.

The overall sum of group and user activity over T1 and T2 is higher than we estimated resulting in using 3.9 PB total compared to 2.0 in our model.

Processing and I/O buffers: This scales with the amount of I/O at a site, which roughly scales with the number of jobs (production, user, group…) running at the site. Our current usage was adjusted to avoid bottlenecks at each site.

Table 4. ATLAS Tier-1 CPU Usage

Tier-1 CPU [kHS06]	***New 2010***	***Scaled by calendar***	***Scaled by live seconds***	***Present***
Reprocessing	15	11	n/a	1.4
Simulation Production	109	82	n/a	57
Simulation reconstruction	16	12	n/a	12
Group+User activities	38	28	n/a	31
Total	***178***	**133**	n/a	102

Comments on Tier-1 CPU

Total Present: This figure entered in the Total Present cell is derived by scaling the WLCG reported accumulated wall clock time (28 million HP06-days) in the August 2010 report.

Only the reprocessing component of T1 CPU usage scales with the calendar.

Simulation production is clearly the dominant activity at Tier 1, but with substantial group and user activity. The proportion of group and user activity at Tier 1 *vs.* Tier 2 reflects the early running where groups and users needed the ESD, which is placed at the T1 centres according to our computing model. We expect this proportion to shift as usage of ESD is largely supplanted by usage of larger derived data sets.

Both a major simulation production and reconstruction campaign and a major reprocessing campaign are being prepared for autumn of this year.

Table 5. ATLAS Tier-1 Tape Usage

Tier-1 Tape [PB]	***New 2010***	***Scaled by calendar***	***Scaled by live seconds***	***Present***	***Comment***
Raw data	2.7	2	1	4.0	Includes cosmics and all RAW and real data formats
Real ESD/AOD/DPD data	3.3	2.5	1		
Cosmics and other data	4.3	3.2	1.4		
Simulated data	7.8	6	2.5	1.6	
Group+User	0.1	0.07	0	~0	
Total	***18***	**13**	***6***	**5.6**	

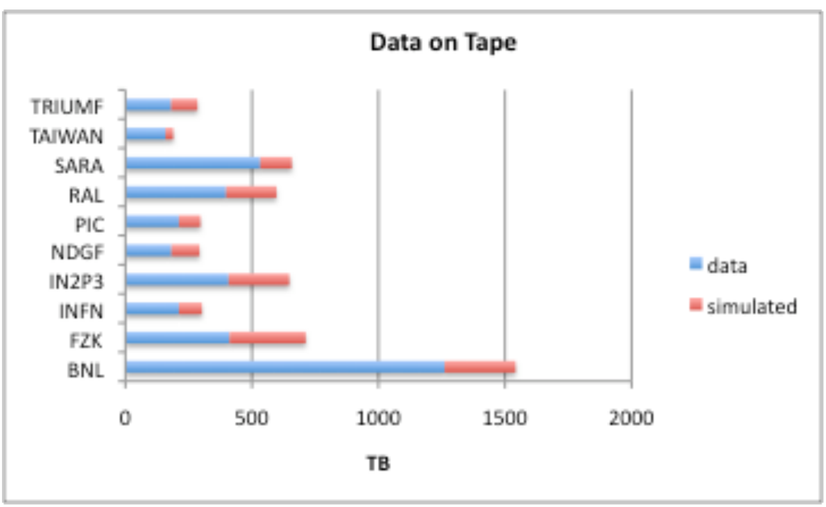


Figure 1. ATLAS accounting chart showing tape usage.

Comments on Tier-1 tape

RAW data, Real ESD/AOD/DPD data, Cosmics and other data: We only have access to the total amount of real data on tape, not the breakdown between cosmics, raw and derived formats. Although the amount of new data should scale with live time, there is a also substantial volume

(2.9 PB) of legacy data from earlier years. (We do not erase or re-use tapes.) The amount of real data from 2010 on tape at the T1s is 1.1 PB. Early on in the run this year, we decided not to put derived data on tape, since it can be regenerated and often we have more than one copy on disk.
Simulated data: This doesn't scale with live time, but is determined by the timing of our simulation campaigns and the number of events simulated in these campaigns. The overall tape volume for simulated data is smaller than our model, despite having simulated more events. We decided to not save any derived data to tape, since it could be easily reconstructed from the simulated HITS (called MC RAW in our computing model documentation). In addition, we save only a fraction of RDO data.

Tier 2

Table 6. ATLAS Disk Usage at Tier-2 Centres

Tier-2 Disk [PB]	*New 2010*	*Scaled by calendar*	*Scaled by live seconds*	*Present*	**Present Primary**	**Comment**
Current RAW data	0.4	0.3	0.13	4.6	1.5	This is the sum of "Current RAW data" and Real (AOD+DPD) data.
Real AOD+DPD data	11.3	8.5	3.6			
Simulated data	7.7	5.8	2.5	3.7	1.2	
Calibration and alignment output	0.12			0.13		
Group data	2.8	2	1	1.4		
User data	1.2	1	0.4	1.5		
Processing and I/O buffers	0.6	0.45	0.2	1		
Total	**24**	**18**	**8**	**12**	**3**	

Comments on Tier-2 disk

Current RAW data and **Real AOD+DPD data:** Our plan was to have a rather large amount of RAW data on disk at Tier-2's, but we found this to be not as useful for physics analysis; consequently, we reduced the amount of raw going to Tier-2.

So far this year, our largest reprocessing campaign processed 168 million events. Our computing model estimates are based on 3 campaigns, with the full data sample near the end of the year. The derived data (ESD, AOD, DPD) on disk will go up dramatically after the next 2 reprocessing campaigns. Note, we have 820 million events now.

Simulated data: We have simulated more data than planned, but most of the simulated data is stored on Tier-1 disk, even though it is produced at Tier-2's. We have also been running our automatic deletion program for the past 2 months. This program deletes data from disk at T2s that has not been used recently and exists at some other site. Thus, for some of our most used data, we have multiple copies of AOD according to our computing model, for other data we have fewer copies than in our computing model.

Group data and User data: We do not expect this component to scale with live time as users and groups keep disks full with the most useful data samples. As mentioned above for T1, the group production is still ramping up.

Processing and I/O buffers: The activities at T2 do not scale with live time. The simulation and analysis jobs are run continuously, with frequent file transfers. The buffers are needed to handle the file transfer load.

Table 7. ATLAS CPU Usage at Tier2

Tier-2 CPU [kHS06]	*New 2010*	*Scaled by calendar*	*Scaled by live seconds*	*Present*
Simulation Production	65		n/a	121
Group activities	38		n/a	24
User activities	123		n/a	
Total	***226***		n/a	145

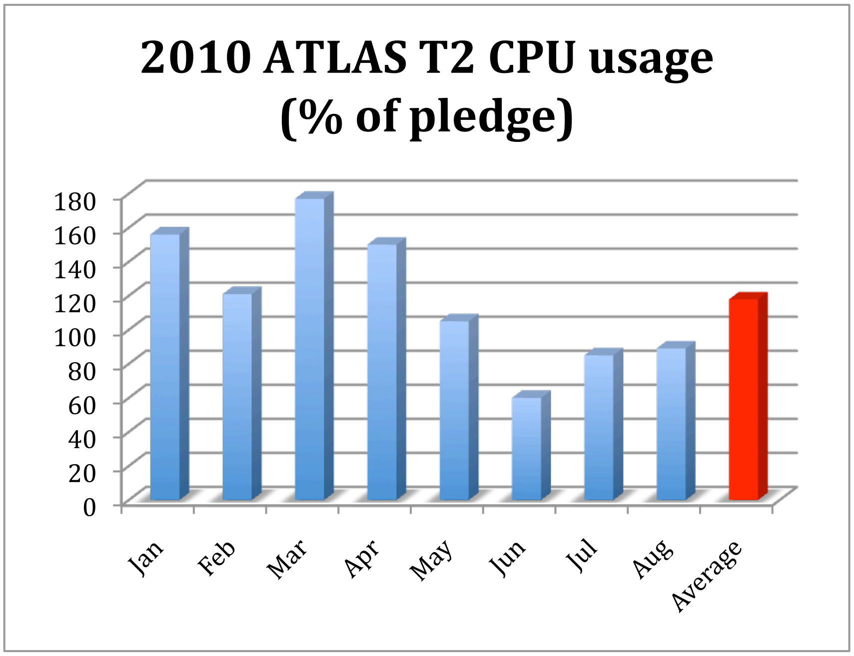


Figure 2. WLCG accounting chart showing CPU usage as a percentage of the pledged resources.

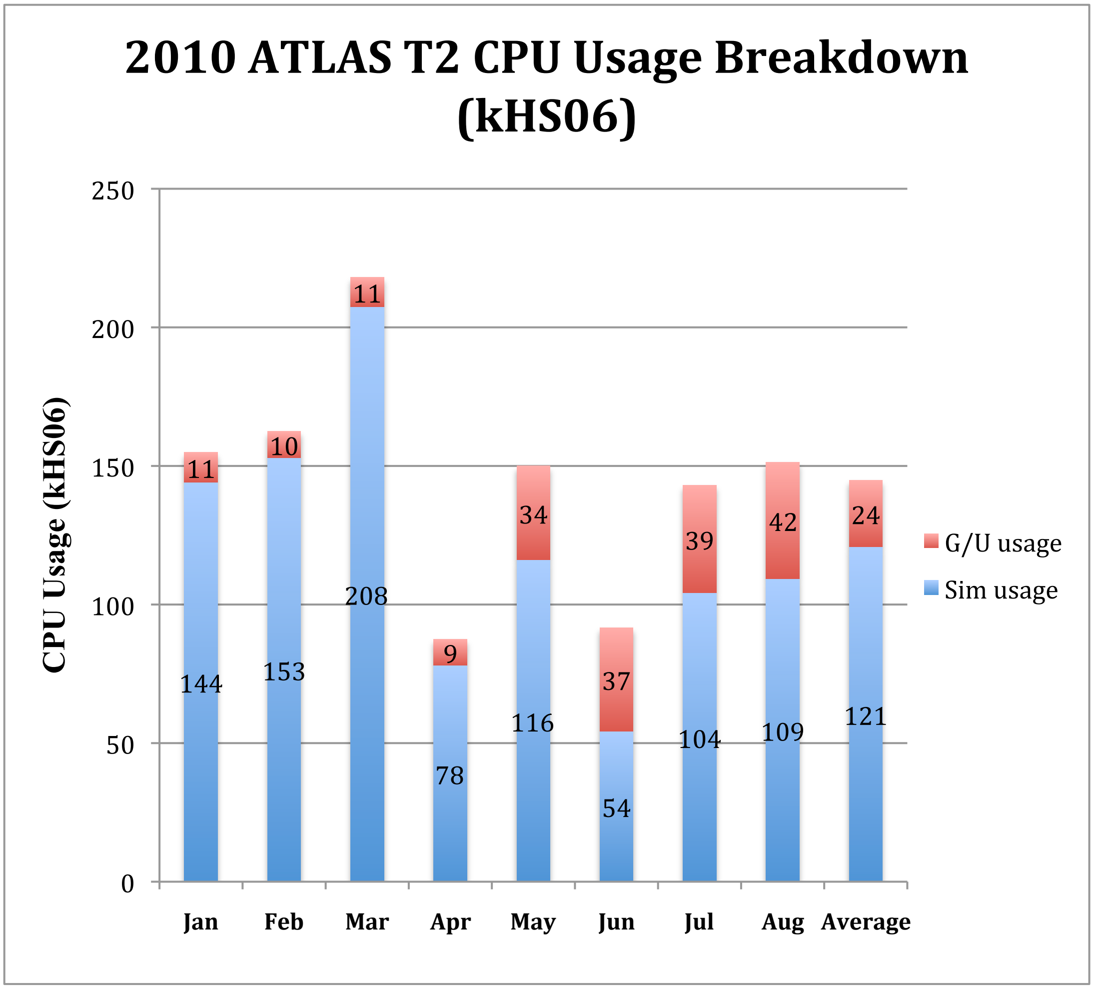


Figure 3. Tier 2 CPU usage breakdown showing simulation production and Group/User activity.

Comments on Tier-2 CPU

To date, CPU usage at T2 has been dominated by simulation; however, we have seen a steady increase in group and user usage. The ATLAS accounting, from the ATLAS production database, shows, that averaged over the year to date, the T2's spent 82% of the CPU usage on simulation and the rest on group and user analysis. We expect group and user usage to increase dramatically as we get larger derived data sets from higher luminosity running and as need for ESD diminishes. As mentioned above, in the early running, groups and users needed the ESD, which is at the T1 centres.

Our accounting at this stage does not separately account for the group and user usage. The group and user activity has been ramping up steeply this year, so averaging over the full year, as in Table 7, does not properly reflect the current usage.

Figure 2 shows the WLCG accounting chart of T2 CPU monthly usage as a percentage of pledged resources. Recall that the 2010 WLCG pledges were expected to be in place on 1 June 2010. During the months preceding June, T2 usage exceeded pledges, and during the subsequent months has been somewhat less than the pledges. The average over the 8 months shown for ATLAS usage is 120% of the pledged resources. This is the number appearing in Table 2. The 145 kHS06 listed in Table 7 represents the average over the 8 months of 2010 and comes from the WLCG accounting report.

CERN: Tier-0 and CAF

CERN CPU: Although the average usage of Tier-0 CPU resources is ~30%, usage is 100% for periods of many hours, or even days, when there is beam. In fact, Tier 0 capacity has become one of the factors that introduces delay in making reconstructed data available; consequently, the installed 28 kHS06 is the minimum needed to keep up during data taking. In addition, 12 kHS06 are used in the CAF for the software build system, operational servers, and the ATLAS Run Time Tester (RTT). Another 23 kHS06 are used for non-automated calibration and other detector studies.

CERN disk: The size of the Tier-0 disk pool is mainly determined by the I/O rate needed for reading, writing, processing, and merging the RAW and derived data. Currently there are 93 servers in the <t0atlas> and <t0merge> pools, leading to a 915 TB disk pool. In addition, the CAF has 1.5 PB of data used for performance studies. This includes 0.8 PB of real raw data and 0.6 of Monte Carlo.

CERN tape: Current tape usage at CERN is shown here in TB:

	AOD	CBNT	DESDM	DPD	ESD	NTUP	Other	RAW
8/31/10	111	289	80	33	1667	254	171	4712

This shows a total of 7.3 PB of data on tape at CERN, dominated by RAW and ESD data.
AOD…Analysis Object Data
CBNT….Combined N-Tuple
DESDM…compact data derived from the ESD
DPD…..Derived physics data
ESD…..Event Summary Data
NTUP….N-tuple used for calibration, alignment and performance studies.

CMS Resource Utilization in 2010

Overview:

The 2010 computing operations in CMS has been reasonably smooth so far. The performance of the machine both in terms of live time and integrated luminosity is smaller than was planned for in 2009. The target after the first month of commissioning was nearly 40% live time. Figure 1 shows the machine live time in 2010 until August, and the target has only been achieved for 3 weeks. The final 6 weeks of the 2010 run should be indicative of 2011 running. The lower live time of the machine has manifested itself as a lower utilization of the Tier-0 processing resources and faster turn around for data reprocessing at the Tier-1s.

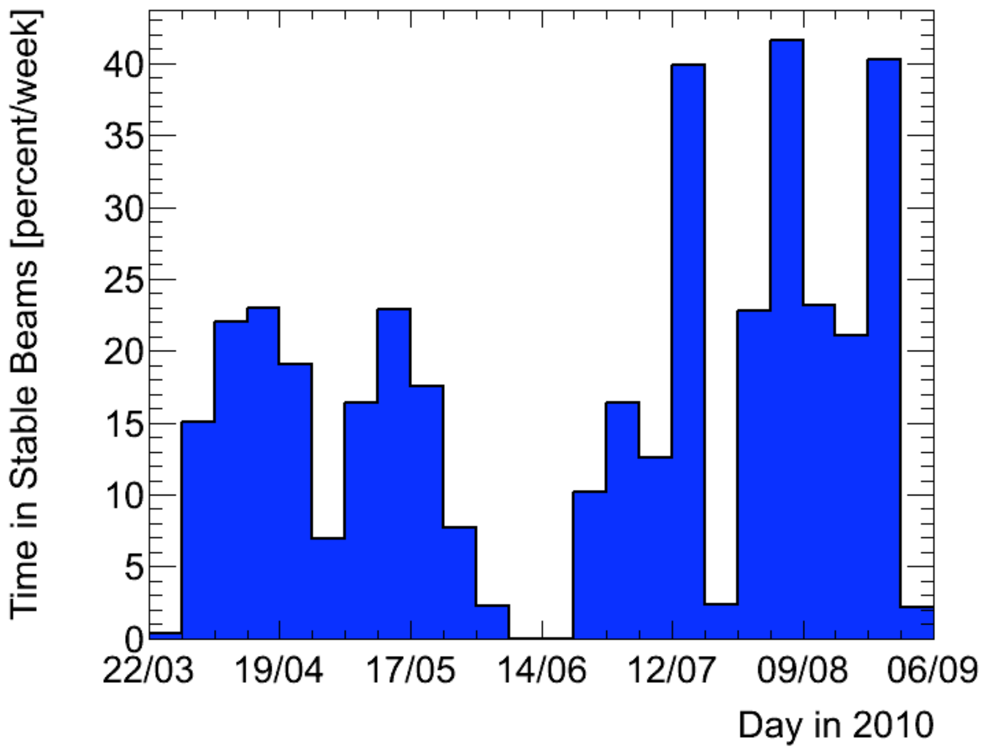


Figure 1: LHC Percentage of time in stable beam during 2010.

While the average utilization was lower that expected, the instantaneous trigger rate was within expectations. During the early beam running CMS was very conservative and operated with largely open triggers. The overall trigger rate was close to the 300Hz design, but CMS recorded many non-collision events. The instantaneous luminosity rose by over 5 orders of magnitude during 2010. By deploying increasingly selective triggers CMS was able to maintain trigger rates close to 300Hz. Toward the end of the summer CMS was averaging over 400Hz to give flexibility in deploying a more selective trigger. The lower luminosity events in 2010 have had a lower reconstruction time than planned for, and the Tier-0 was able to accommodate the higher rate.

The prompt reconstruction and storage systems at the Tier-0 have performed within the design parameters. The Express stream is approximately 10% of the data and is reconstructed within an hour to monitor the performance of the detector. The Tier-0 has nearly always met the reconstruction targets. The prompt reconstruction software has been extremely stable only a few failed reconstruction jobs in 2010.

The transfers to the Tier-1 sites have been smooth and reliable. The CMS raw event size for data has been much smaller than planned for due to the simpler events at lower luminosity, but also the detector has been quieter than the initial Computing TDR estimates. The reconstruction size is very close to the planning estimates. Figure 2 shows the rate from CERN out to the Tier-1s at the end of August. The rate at this luminosity reached 1GB/s from CERN, which is very close to the expected peak target. The balance of bandwidth utilization across Tier-1s roughly matches the fraction of resources indicating each site is contributing.

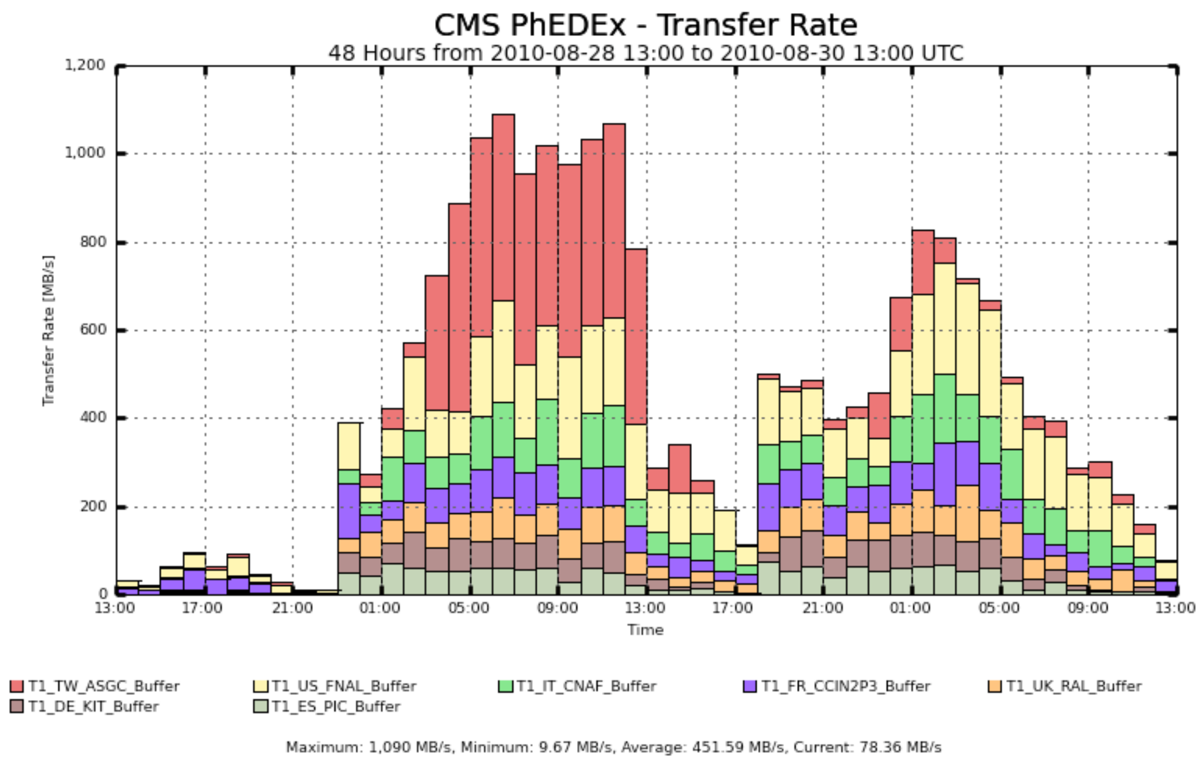


Figure 2: CMS Transfer rate CERN to Tier-1s during the end of August, 2010

From the start of running up to ICHEP CMS had many software and calibration improvements. The frequency of data reprocessing was higher than planned for, but could be accommodated with the smaller volume of data. Ten partial or complete processing passes over the data were performed in the first 6 months of running. The rate of reprocessing has diminished in the late summer, and two complete reprocessing passes are proposed for the final weeks of running and the preparation for winter conferences.

The transfers to the Tier-2 performed well with the first blocks of data routinely accessible for analysis users within hours of initial distribution. CMS commissioned additional Tier-2 to Tier-2 transfer links during 2010. These are proving themselves to be very useful in replicating data to Tier-2s. The number of users actively submitting analysis jobs per week to the grid resources for the last year is shown in Figure 3. The number of jobs submitted per day by the users is shown in Figure 4. The rate of analysis submission reaches the Computing TDR estimate shortly after the run starts and continues throughout the run. The high level of CMS analysis activity is also visible in the high fraction of Tier-2 resources used.

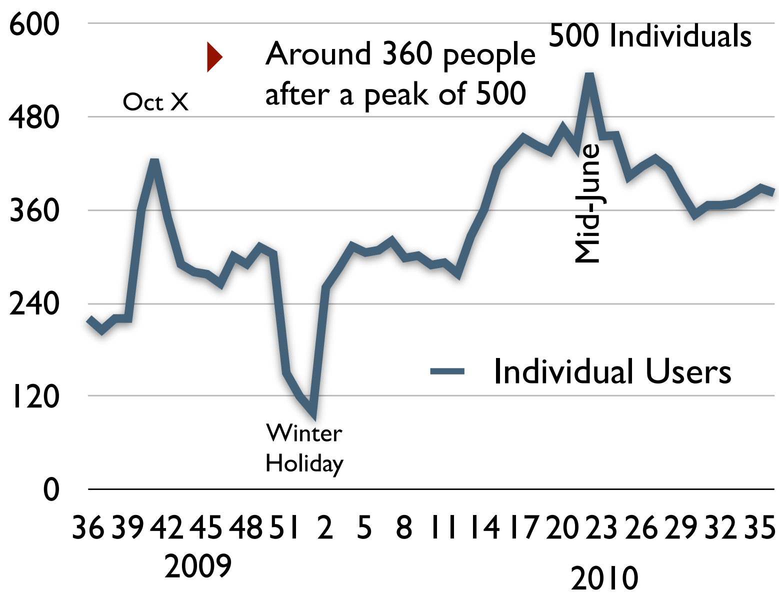


Figure 3: Number of CMS users submitting analysis jobs per week.

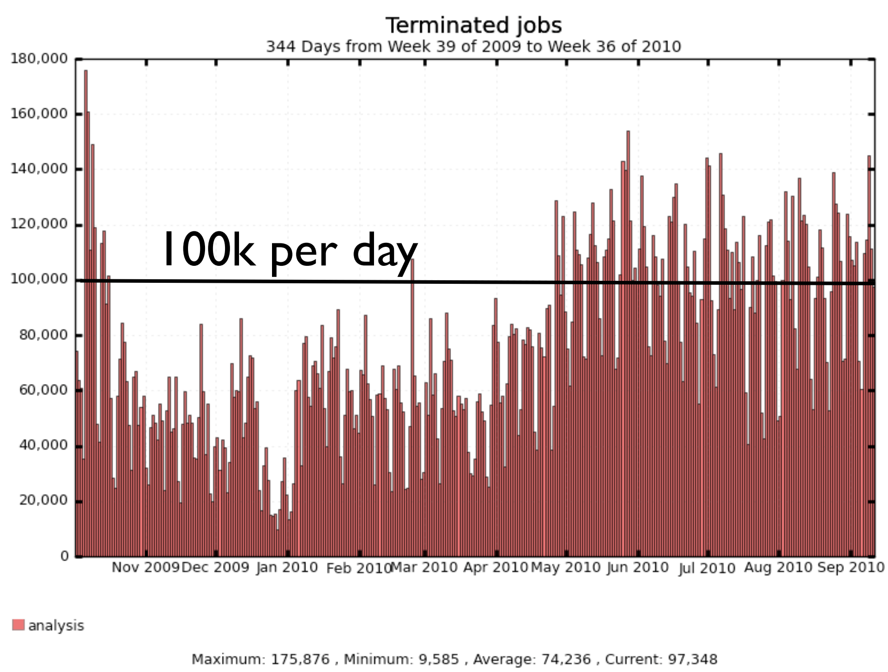


Figure 4: Number of jobs submitted by analysis users in 2010.

CPU:
The CPU utilization in CMS is described in the common utilization document. Overall the utilization of CERN and the Tier-1s is lower than planned. This is related to the machine live time and the complexity of the 2010 data. The CPU efficiency for organized processing at CERN and the Tier-1s is on average 75%. This is close to the computing model expectations. The next generation of CMSSW (CMSSW_3_8) that was recently deployed on the Tier-0 is expected to improve the CPU efficiency for organized processing.

The CPU utilization of the Tier-2s is high and close to the pledged values. The Tier-2s have been heavily used for simulation and analysis. The heavy use of the Tier-2s is an indication of the success of moving people to use distributed computing resources. The high level of usage given the low amount of accumulated luminosity is somewhat concerning, but may scale with experiment enthusiasm more than total events collected. The CPU efficiency at the start of data taking dropped as the fraction of chaotic analysis increased. The deployment of CMSSW_3_8 for analysis is expected to improve the CPU efficiency.

Storage:
In 2010 CMS has collected approximately 700M collision events. Especially in the early running there was a high fraction of test triggers and zero bias events, which factor into the RAW and RECO dataset sizes. During the same time CMS has produced about 760M simulated events. CMS has requests to be completed by the end of 2010 for an additional 950M simulated events. Table 1 shows the total volume of RAW events and RECO events collected in 2010 data and the volume of new simulated data produced in 2010. This does not include commissioning data collected in previous years or simulated events from 2009. For RECO this does not include a few processing passes that were produced but later declared deprecated and scheduled for cleanup.

Tier	Size
RAW	455TB
RECO	670TB
Simulation (RAW + RECO)	950TB

Table 1: New data and simulation produced in CMS during 2010

Disk:
The disk space used by CMS at the end of August is shown in Table 2 for CERN, the Tier-1s and the Tier-2s. The percentage of disk space used at CERN is lower that the other sites in part because it is dominated by new data produced in 2010, while the other sites have a reasonable fraction of simulated events produced in 2009. The disk

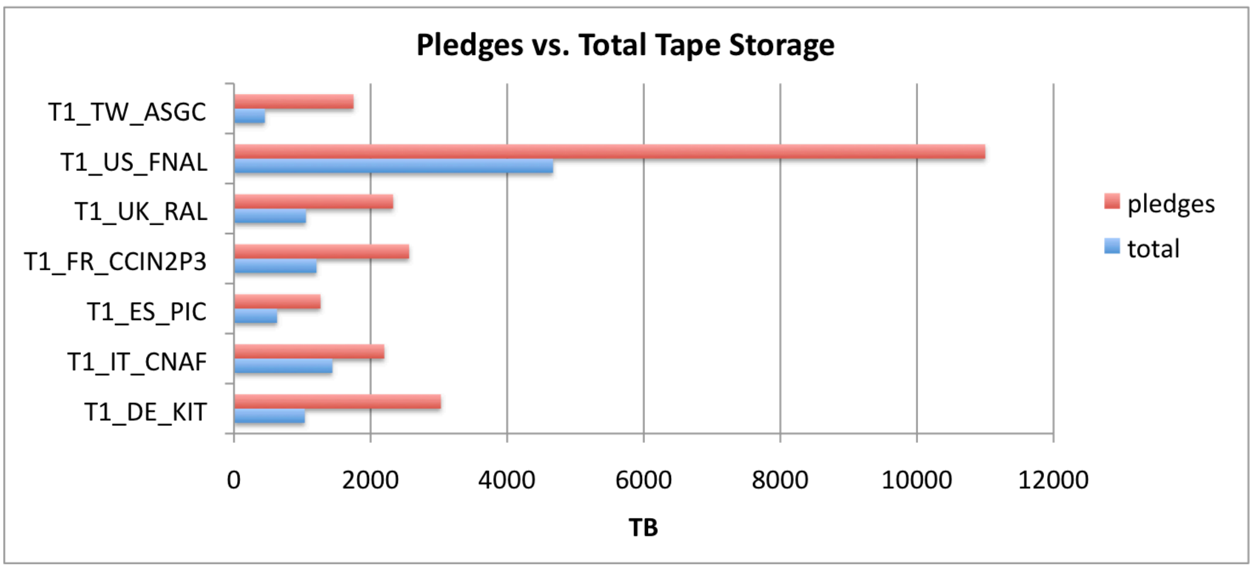
fill will be assessed during the higher luminosity runs in the fall of 2010 and the Heavy Ion run.

Site	Space Used	Space Available
CERN	2.5PB	4.4PB
Tier-1s Only	8.3PB	8.7PB
Tier-2s Only	7.5PB	13.6PB (9.2PB requested)

Table 2: Disk Space used by CMS at Computing Centers at the end of August, 2010.

Tape:

The utilization of tapes at the Tier-1s is shown in Figure 5. September is 50% of the resource year and CMS is in good shape to finish the 2010 year.



LHCb Grid Computing Resources usage in 2010

V2, 04/10/2010

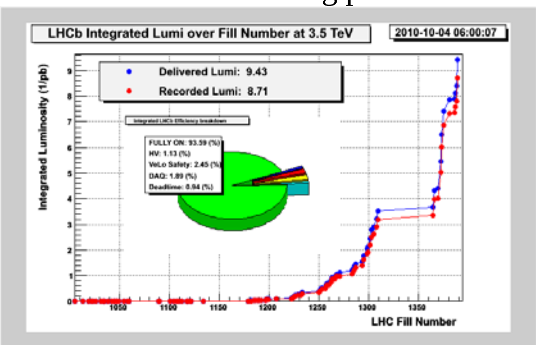
At the Computing RRB in April 2009, LHCb presented some resources requests for the 2009 running, in fact for a period covering April 2009 to April 2011. This request was then updated and a final version was approved by the C-RSG in June 2009:

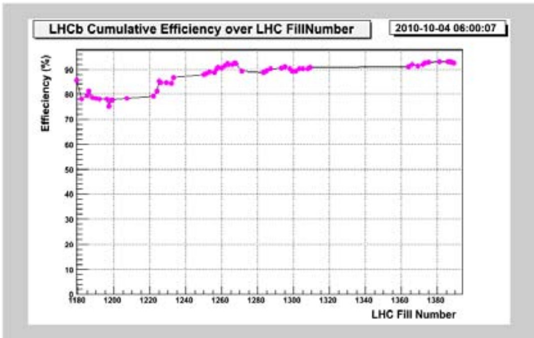
Date	Site	kHS06	Disk (TB)	Tape (TB)
Oct'09	CERN	15	720	1000
	Tier-1	31	1740	1000
	Tier-2	31	20	0
Apr'10	CERN	21	1290	1500
	Tier-1	41	3290	1800
	Tier-2	36	20	0
Oct'10	CERN	23	1290	1800
	Tier-1	44	3290	2400
	Tier-2	38	20	0

Table 1: CPU Power, disk and tape storage needed in place to meet LHCb requirements for the 6 month period commencing (a) October 2009, (b) April 2010 and (c) October 2010, as best estimate end of June 2009.

Data collected in 2010

Since March 30th 2010, LHCb has been collecting real data with a quite high efficiency as shown on the following plots:





LHCb collected 116.6 TiB of raw data from March 30th until end September 2010. The RAW data were distributed immediately to the 6 Tier1s using the following share[1]:

Raw data		
SE	**Size (TiB)**	**Nb of files**
CERN-RAW	86.9	74981
CERN-RDST	37.8	34322
CERN	**124.7**	**109303**
CNAF-RAW	14.3	12871
GRIDKA-RAW	21.1	18963
IN2P3-RAW	26.7	24237
NIKHEF-RAW	26.3	23786
PIC-RAW	6.6	6037
RAL-RAW	21.6	19311
Tier1s	**116.6**	**105205**

Table 2: RAW data collected by LHCb in 2010 until end September

1 At CERN, the 124.7 TiB include data taken for detector calibration that was not distributed.

The following figure shows the number of concurrent jobs running on the Grid since March 30th.

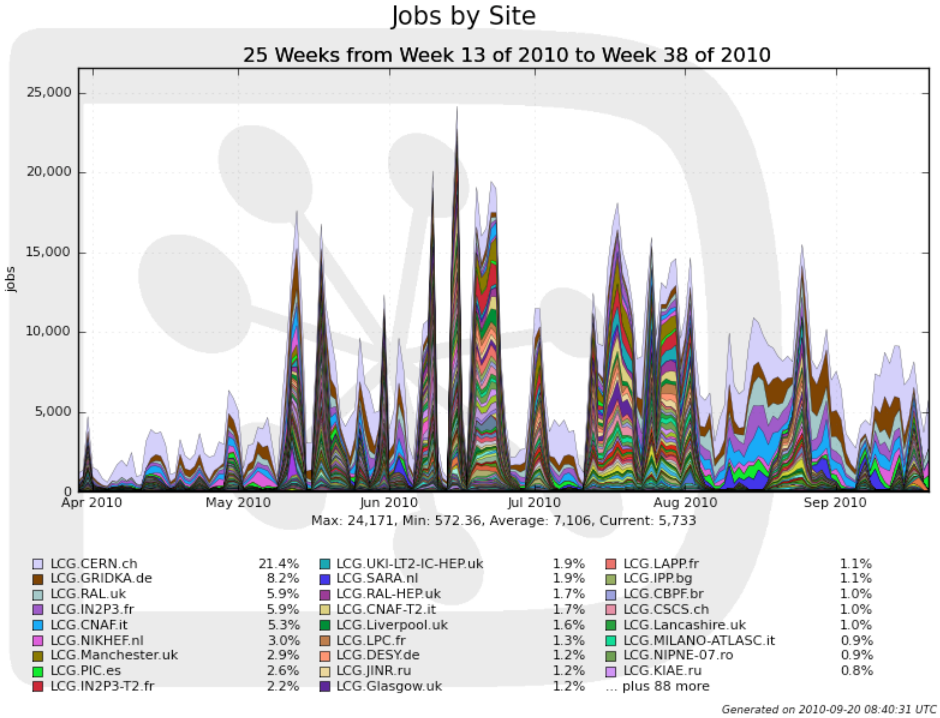


Figure 1: Number of running jobs by LHCb, split by site

One can see than 114 sites have been used during this period. There have been peaks of close to 25,000 jobs running simultaneously in June during a period of intensive Monte-Carlo simulation.

Restricting to Tier1s, the following plot shows the number of jobs completing every hour on the same period (note this doesn't take into account the duration of jobs):

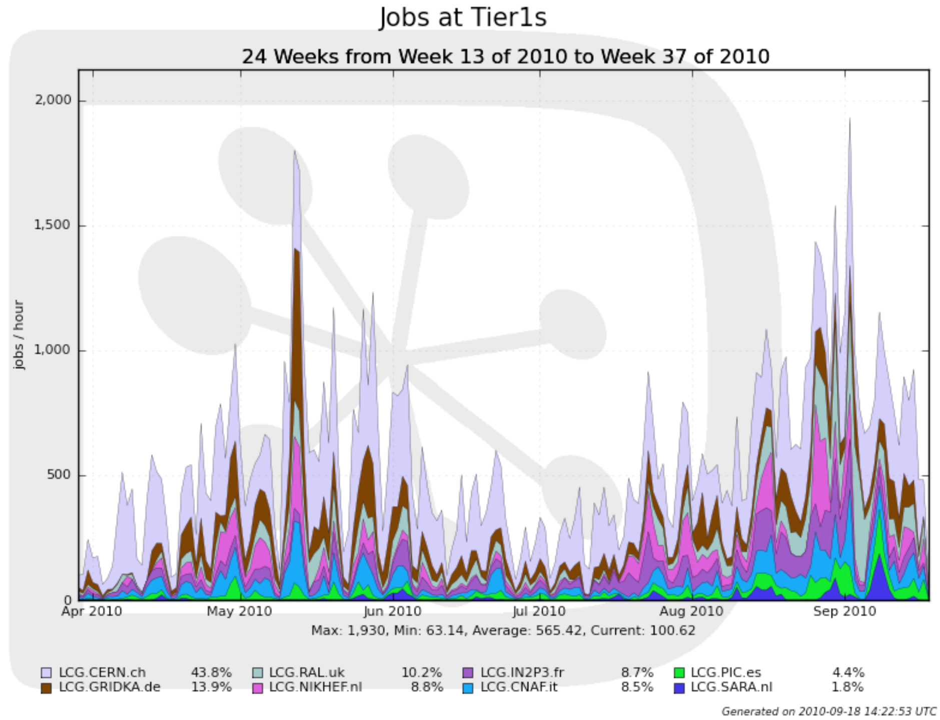


Figure 2: Number of jobs completing per hour at Tier1s

In order to adapt the Computing Model at best with the hardware configuration of the NL-T1, reconstruction and stripping jobs are assigned to SARA while analysis is assigned to NIKHEF. This is why the plot shows 8 Tier1s instead of 7.

No simulation job has been running on Tier1s except during a short period in May as shown on the following graph. The unit is in CPU.hours used every second. An average of 0.53 indicates on average 1900 jobs running on Tier1s.

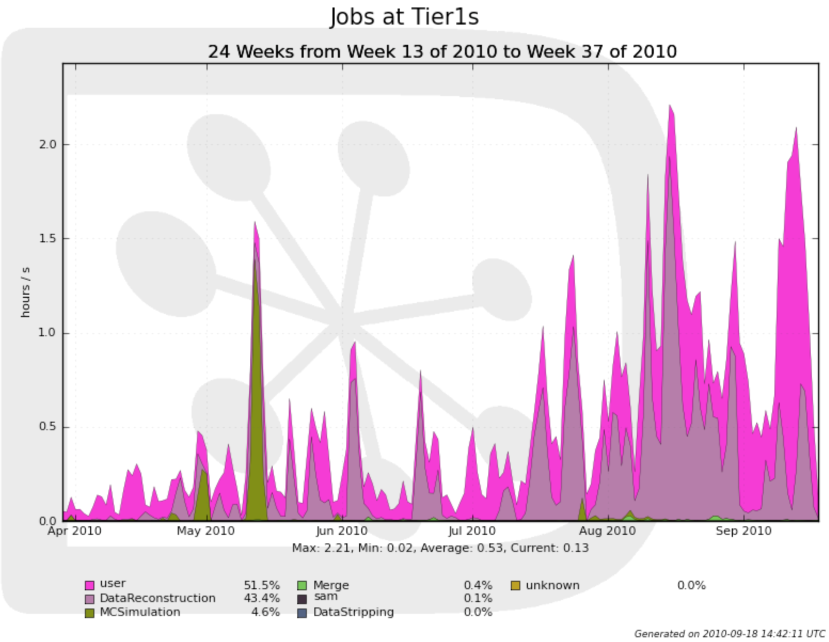


Figure 3: Unnormalised CPU used at Tier1s (in CPU.hours per s)

User jobs represent over 51.5% of the CPU usage at Tier1s, the rest being production activities. One clearly sees on the graph the influence of increasing luminosity on the CPU utilization at Tier1s, in particular for user jobs as shown below:

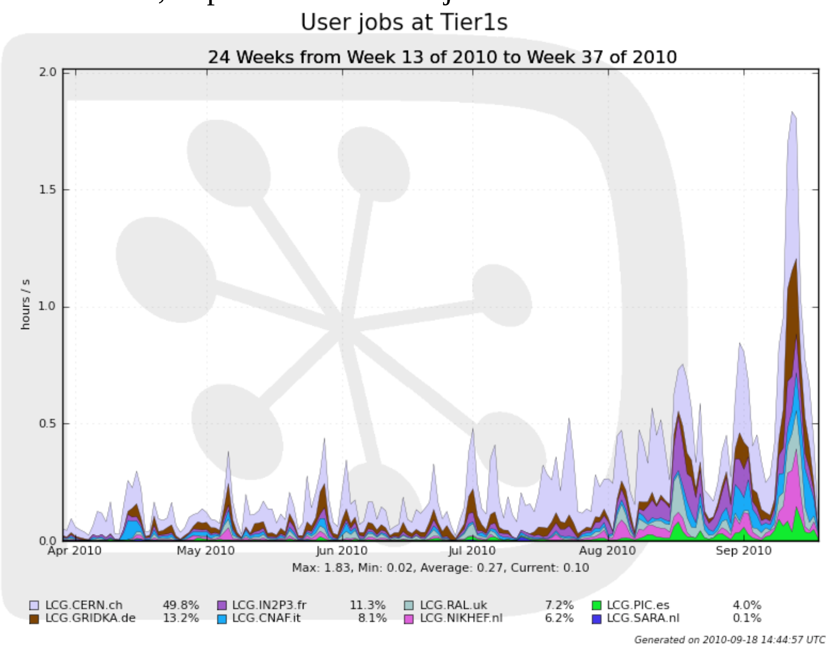


Figure 4: CPU used by user jobs at Tier1s (in CPU.hours per s)

It is worth noticing that the fraction of analysis jobs that have been running at CERN in 2010 is 50%, the share not being fixed by LHCb but by the relative availability of CPU slots at sites (through the WLCG WMS).

The evolution of total CPU usage per country (including all sites) as a function of time is shown in the graph below:

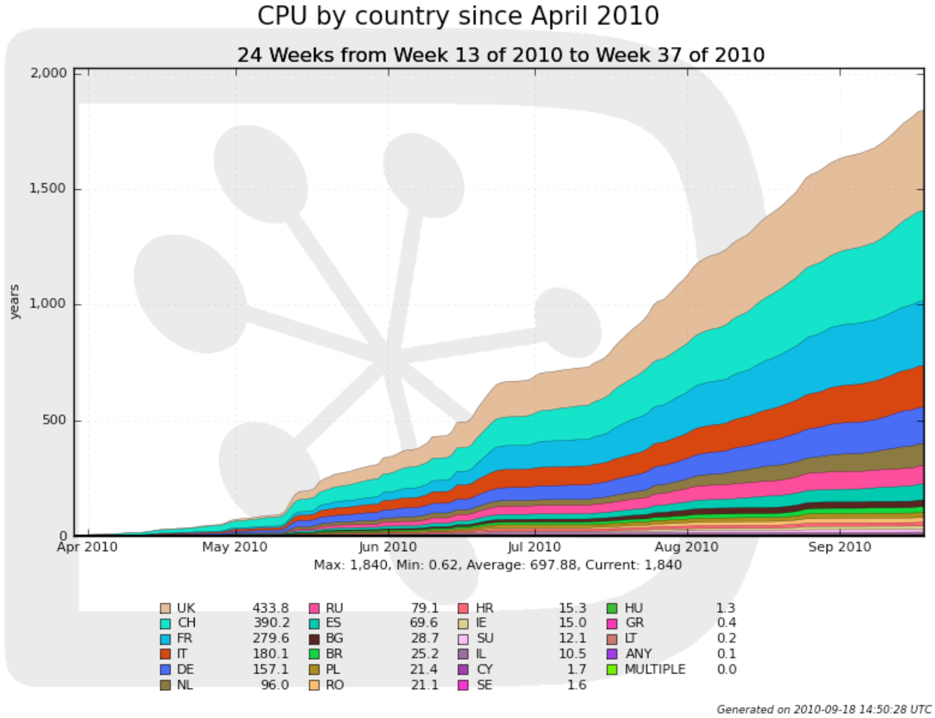


Figure 5: Unnormalised CPU used since April 2010 per country (in CPU.years)

The unit used here is CPU.years (not normalised as the information on the individual Worker Node normalization in HS06 is not available).
Restricting to CERN and Tier1s gives the following share (note that about half the CPU time was consumed at CERN+Tier1s):

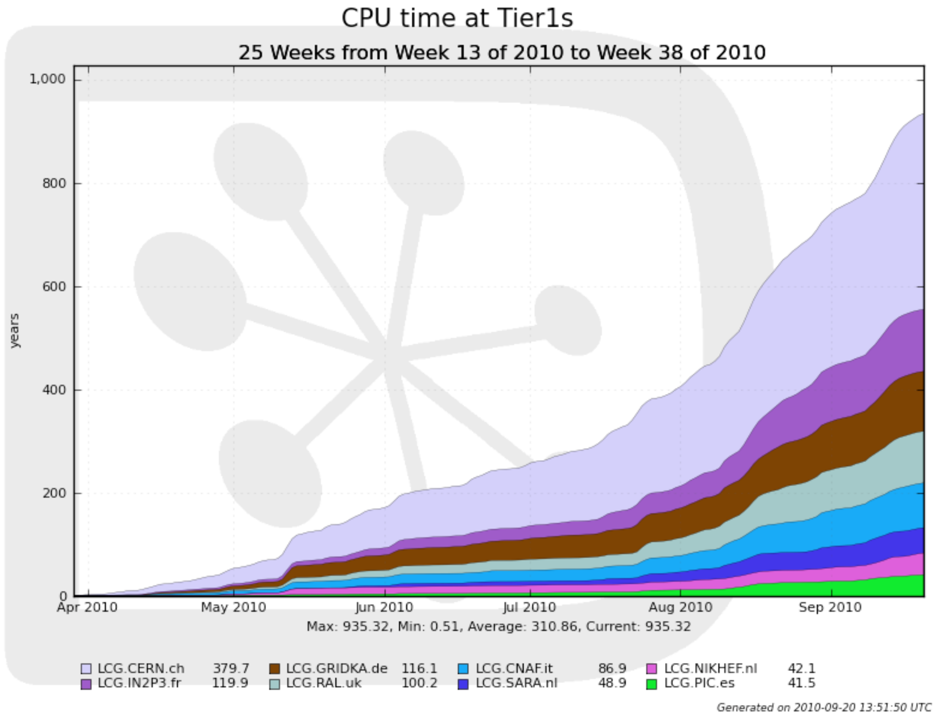


Figure 6: Unnormalised CPU used since April 2010 at Tier1s (in CPU.years)

The reconstruction application had to be adapted to the unforeseen running conditions of the LHC: instead of 0.4 proton-proton visible collisions on average per beam crossing, and due to a higher intensity per bunch than nominally planned, the data collected in 2010 had on average 1.5 visible collisions per crossing, with peaks exceeding 2. The very high occupancy in detectors required special measures to be taken in the pattern recognition and particle identification algorithms. These conditions also had to be taken into account in the physics event preselection algorithms (a.k.a. stripping) that was taking a prohibitive amount of CPU. All these adjustments had to be performed during data taking without much possibility to anticipate the plans form the machine side as they were continuously changing.
The following graph shows the CPU used per event for reconstructing events as a function of the average number of collisions per crossing (μ). The set of points corresponds to different trigger settings. Extrapolating to μ=0.4 (nominal value for LHCb), the reconstruction application is using 12 HS06.s as used for the resources estimates. However the new LHC conditions increase this CPU need by a factor 2 on average.

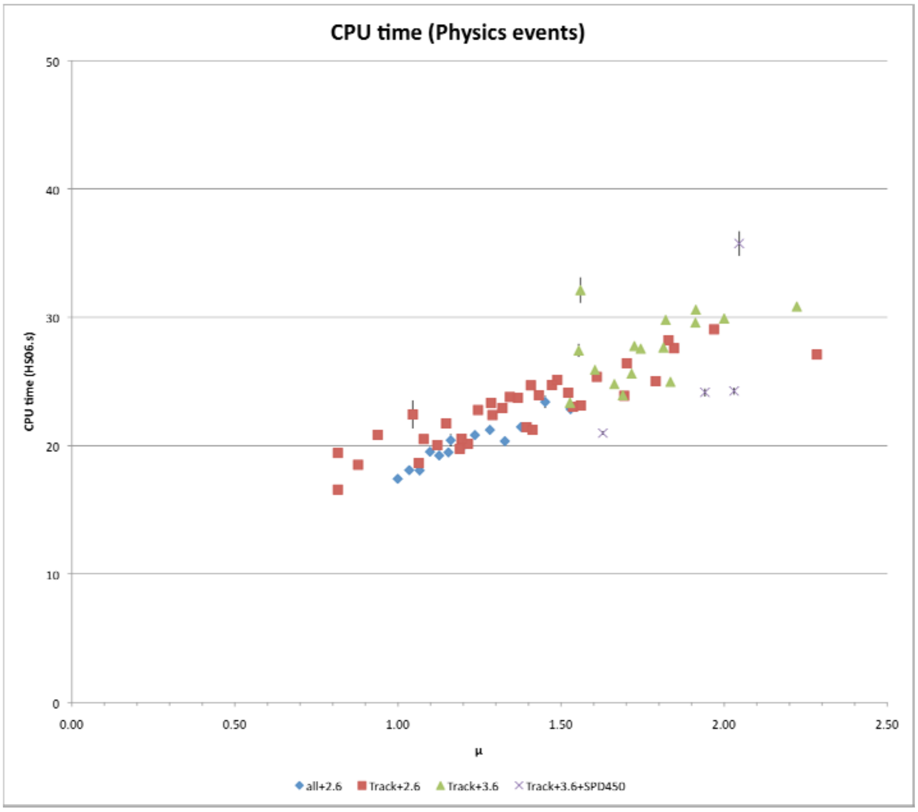


Figure 7: CPU used for reconstructing one event as a function of μ, the average number of collisions per crossing (in HS06.s)

The pattern of CPU usage for reconstruction jobs is shown on the following graph. Besides a continuous reconstruction of incoming data, LHCb has performed 4 partial reprocessings of the existing data at various stages. These were driven by improvements in the reconstruction and stripping applications as well as by a better understanding, alignment and calibration of the detector.

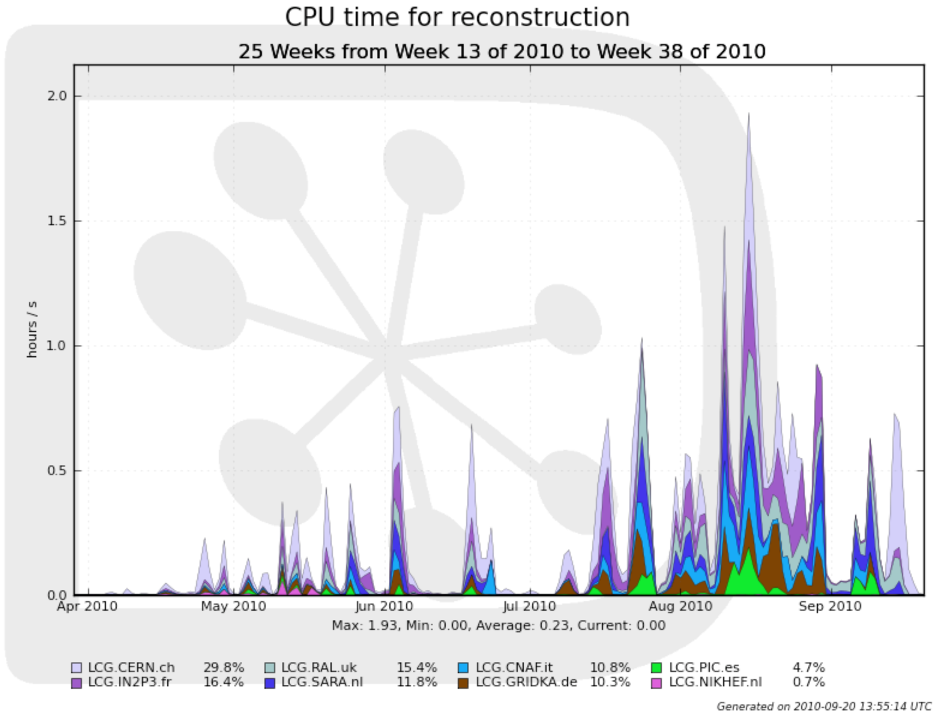


Figure 8: CPU used by reconstruction jobs since April 2010 (in CPU.hours per s)

One can clearly see on the graph the influence of the running pattern of the LHC: during a long period physics data taking was concentrated during week-ends, therefore the weekly pattern of peaks.

From the EGEE accounting portal, the following CPU resources have been used in 2010 for the first 9 months

Sites	<CPU> (HS06)	Requests
CERN	3.2	21
Tier1s	10.3	41
Tier2s	11.5	36
Others	1.2	
Total	**26.2**	**98**

As can be seen from the graphs above the CPU usage is by far not constant. The table above gives an average over the full year. However if one restricts to the last 3 months (July-September) the average power used is:

Sites	<CPU> (HS06)
CERN	4.9
Tier1s	21.0
Tier2s	18.2
Others	1.5
Total	**45.6**

It is clear that the usage is increasing when the luminosity of the LHC increases. As the current requests are valid until April 2011, and CPU usage is by nature not uniform in time, the reserve of a factor 2 doesn't look overwhelming.

Disk usage
During the long machine development period between August and September 2010, a full re-processing of the 2010 data has taken place. At this time, two processing passes are available for physics analysis on disk at CERN and Tier1s (Reco04 and Reco05) on which different stripping strategies have been applied depending on the LHC conditions.
The disk occupancy after this reprocessing was as follows:

Beam3500GeV-VeloClosed-MagDown	**All**	**CERN**	**Replicas**
Reco04-Stripping05	40.6	6.3	6.4
Reco04-Stripping05-PrescaledMinBias	37.3	5.4	6.9
Reco04-Stripping07	391.5	58.2	6.7
Reco05-Stripping09-Merged	322.4	55.4	5.8
Reco05-Stripping09-Prescaled-Merged	39.6	8.2	4.8
Total	**831.4**	**133.6**	**6.2**
Beam3500GeV-VeloClosed-MagUp	**All**	**CERN**	**Replicas**
Reco04-Stripping05	32.0	5.0	6.4
Reco04-Stripping05-PrescaledMinBias	23.8	3.5	6.9
Reco04-Stripping07	338.2	50.2	6.7
Reco05-Stripping09-Merged	163.5	29.5	5.5
Reco05-Stripping09-Prescaled-Merged	105.0	33.9	3.1
Total	**662.5**	**122.1**	**5.4**
Grand total	**1493.9**	**255.7**	**5.8**

The average replication factor is 5.8 and not 7 due to disk storage being full already at several sites. This occupancy is very large compared to the requested disk space for 2010 (350 TB at CERN and 310 TB at each Tier1s for real data DST, hence 1860 TB in total).
This is due several factors:

- A larger event size (140 kB compared to 80 kB), mainly due to the larger average number of collisions (1.5 instead of 0.4)

- The high retention rate of the stripping that has two main causes: the need to apply loose cuts for doing the first analysis and understanding the data; the LHC running conditions.

The following corrective actions have been taken on this data sample in order to be able to have enough space on disk for the still-to-come data:

- Remove all Reco04 DSTs from CERN

- Keep only 3 replicas outside CERN of all datasets

All measures are being taken for tuning the stripping preselections in order to achieve the 10% retention foreseen or below. In cases when the number of events does not allow, the data will be written out in the form of μDST that uses 10 times less space than the DST (to the price of not containing the RAW data and not allowing refitting of tracks, vertices etc…). The yearly full reprocessing should take place in November-December and follow those requirements.
Concerning simulated data, a large number of signal and background events has been generated following the requests of physics groups: the MC09 sample was used for preparing the 2010 data taking while the MC2010 sample was simulated using the beam

conditions as available actually in 2010. The DSTs were distributed according to the Computing Model, i.e. one replica at CERN, one replica with tape backend (MC_M-DST) at one Tier1 and another replica on disk only (MC_DST) at a second Tier1. Finally a legacy simulation dataset (DC06) still is available at CERN only and is going to be soon deleted.
The total usage for simulated datasets is as follows:

Storage	MC09 (TiB)	2010 (TiB)	DC06 (TiB)	Total
CERN_MC_M-DST	237.2	109.3	75.0	421.5
CNAF_MC-DST	39.5	15.1		109.1
CNAF_MC_M-DST	35.4	19.1		
GRIDKA_MC-DST	44.9	19.7		121.9
GRIDKA_MC_M-DST	38.5	18.8		
IN2P3_MC-DST	51.1	19.9		145
IN2P3_MC_M-DST	55.2	18.8		
NIKHEF_MC-DST	40.4	20.1		112.5
NIKHEF_MC_M-DST	33.1	18.9		
PIC_MC-DST	6.9	13.8		45.6
PIC_MC_M-DST	10.7	14.2		
RAL_MC-DST	52.7	20		143.7
RAL_MC_M-DST	51.7	19.3		
Total	**697.3**	**327.0**	**75.0**	**1099.3**
Replication rate	**2.9**	**3.0**	**1.0**	**2.6**

As far as the total disk usage is concerned, one should consider not only the space for DSTs for real and simulated data, but also for user space as well as the disk space for keeping ~10% of the RAW and RDST files on disk at CERN (for detector and application studies).
One should also consider that 70 TB at each Tier1 was requested as cache to the T1D0 storage. The breakdown of disk space usage (TxD1) is therefore as follows:

	CERN	Tier1s
RAW	37.8	-
SDST	32.7	-
DST	255	1238
MC_DST	421	678
Users	104	148
Total	**850.5**	**2064**
Requests	*1220*	*2870*

Considering that we need space for the winter reprocessing as well as for the data that remains to be taken in 2010 and the data that will be taken in 2011 before the pledges are made available, we shall have to reduce the number of replicas of certain datasets.

Tape usage

Tape is not only used for RAW data custodial storage, but also for SDST, DST and MC-DST storage. The space used by all those datasets are as follows:

Site	Size (TiB)	Files
CERN	**662.9**	**427242**
CNAF	113.9	64398
GRIDKA	125.9	71034
IN2P3	156.8	90080
SARA	119.3	68217
PIC	61.9	33046
RAL	140.2	73572
Total Tier1s	**718.0**	**400347**

These have to be added to the RAW data as specified earlier, which gives a total of 788 TiB at CERN (for 1500 TiB pledged) and 835 TiB at Tier1s (for 2800 pledged). The difference between the pledged and used resources comes simply from the shorter running time of the LHC.

Conclusions

Despite continuously changing beam conditions, LHCb managed to adapts its trigger and its reconstruction, stripping and analysis applications in order to fit in the envelope of Computing resources that have been allocated. After the December reprocessing, LHCb should be in a better situation for evaluating the needs for 2011 in light of these new highly demanding running conditions. If necessary LHCb will discuss with Tier1s on the possibility to slightly increase the disk capacity, or reshuffle the foreseen capacity between SRM spaces, due to the larger event size.

Experiment Tier-2 CPU Reports: ALICE Reports

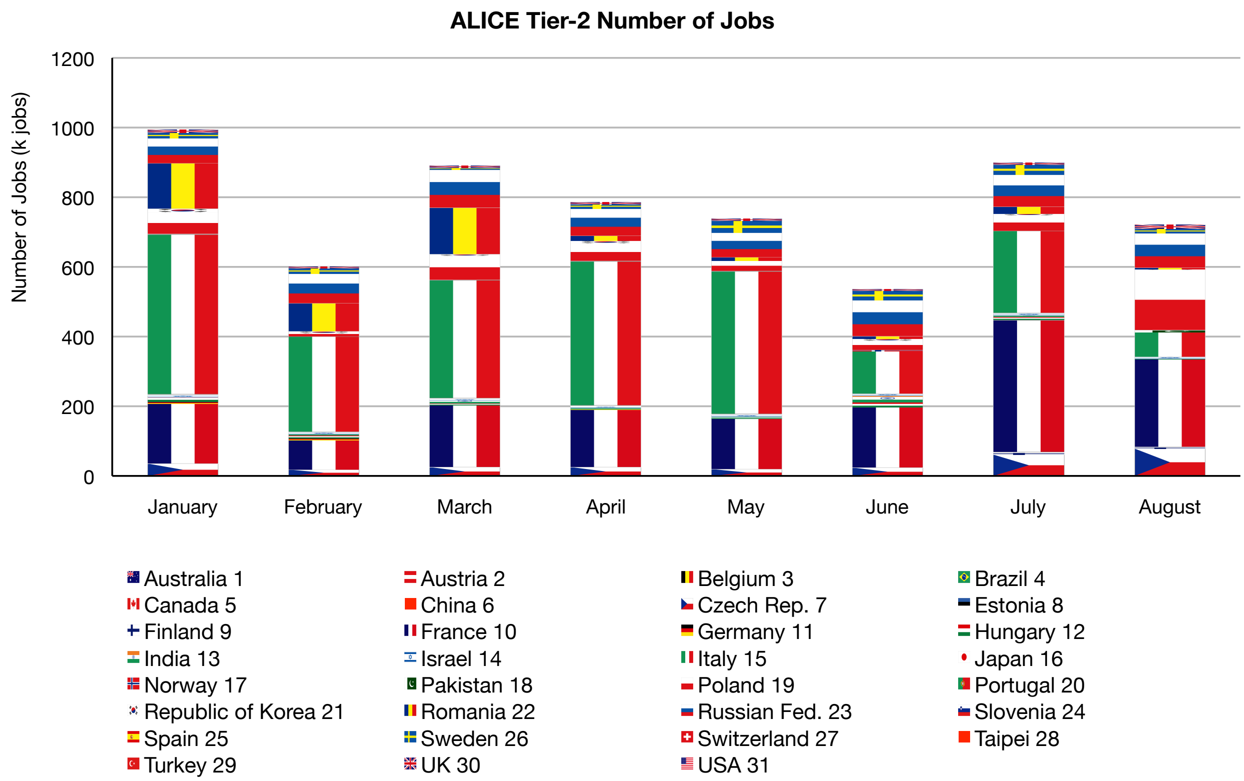


Figure 1: ALICE Jobs Executed on Tier-2 sites during 2010

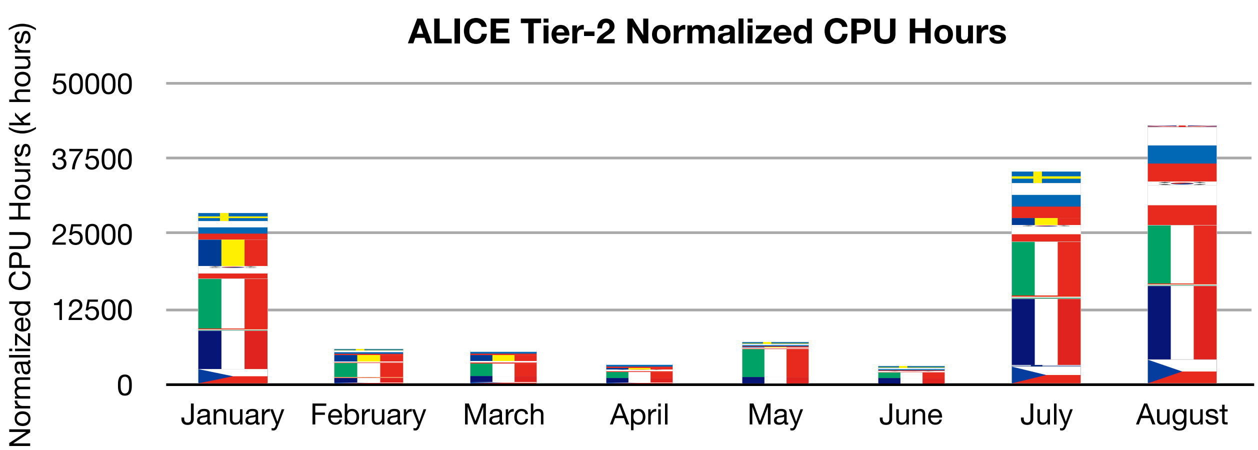


Figure 2: ALICE Normalized CPU Hours at Tier-2 sites in 2010

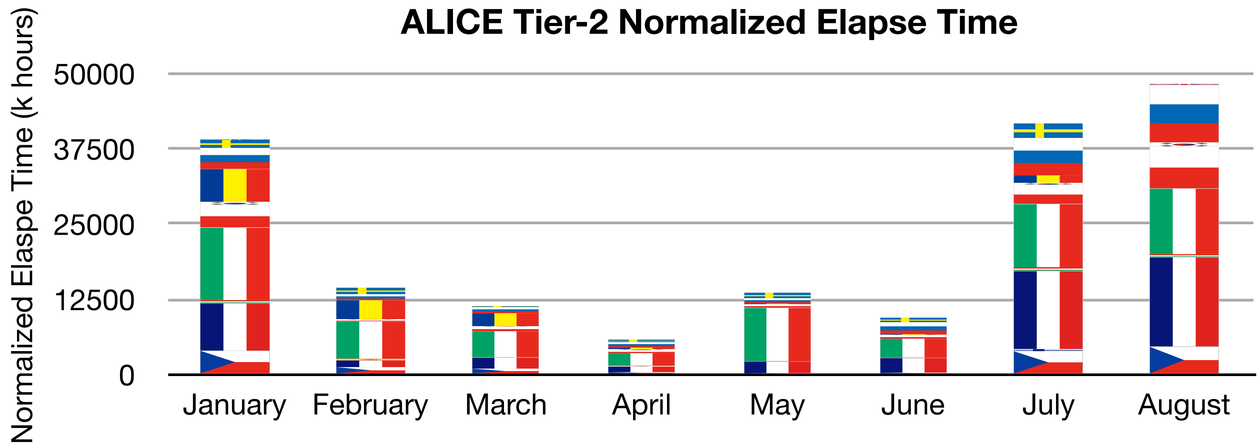


Figure 3: ALICE Normalized Wall Clock Time Used at Tier-2 sites in 2010

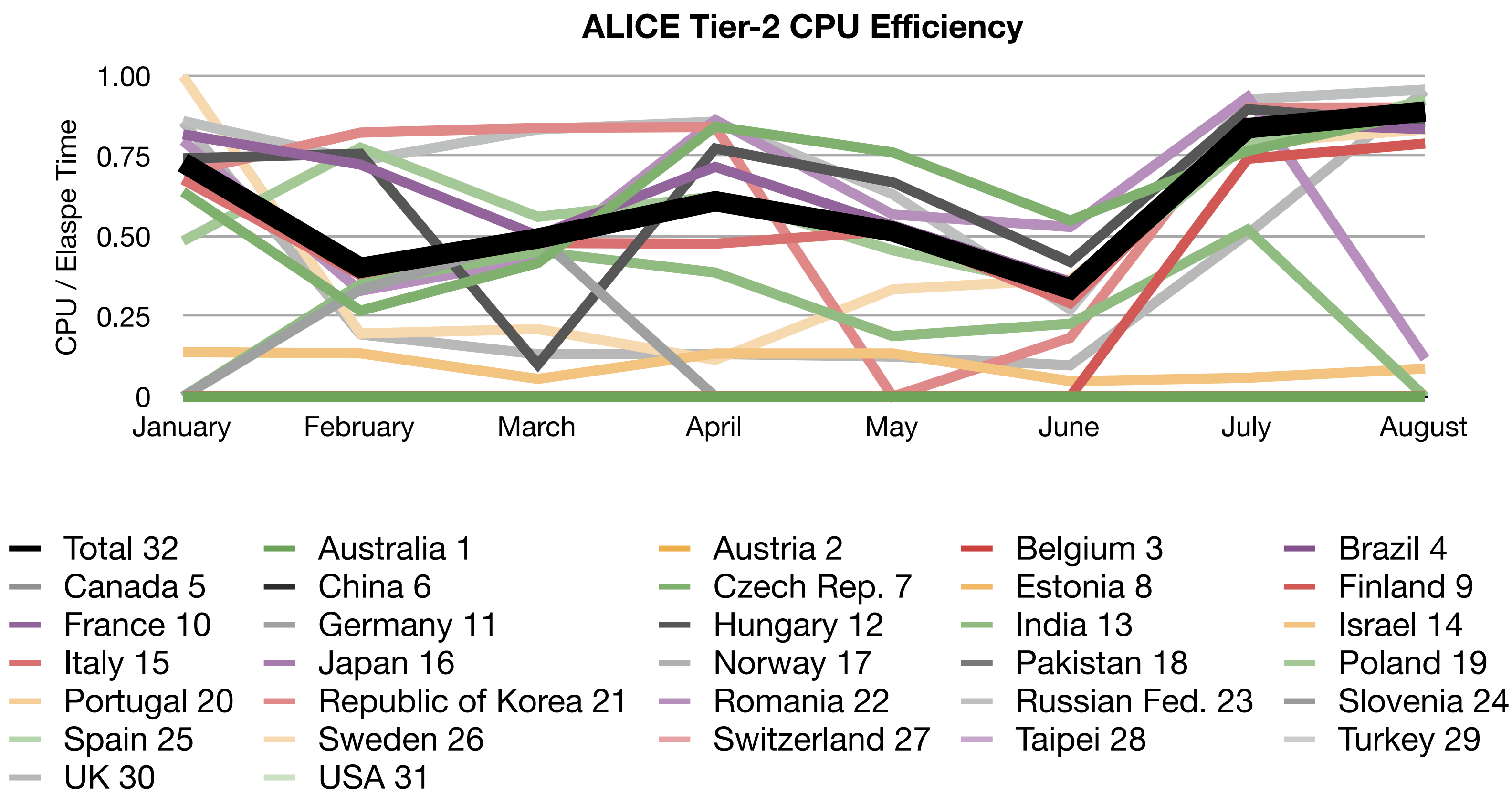


Figure 4: ALICE CPU Efficiency at Tier-2 sites

ATLAS Reports:

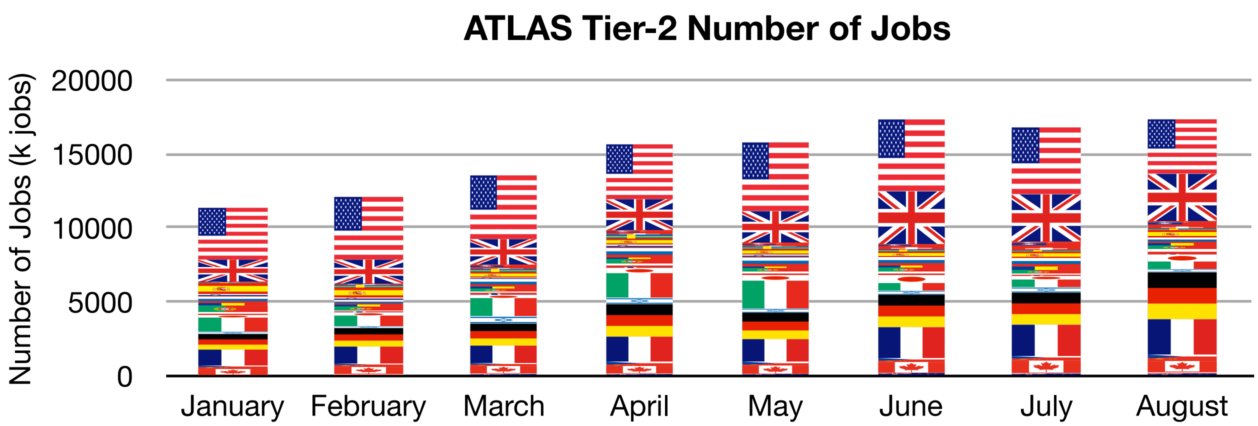


Figure 5: ATLAS Jobs Executed on Tier-2 sites in 2010

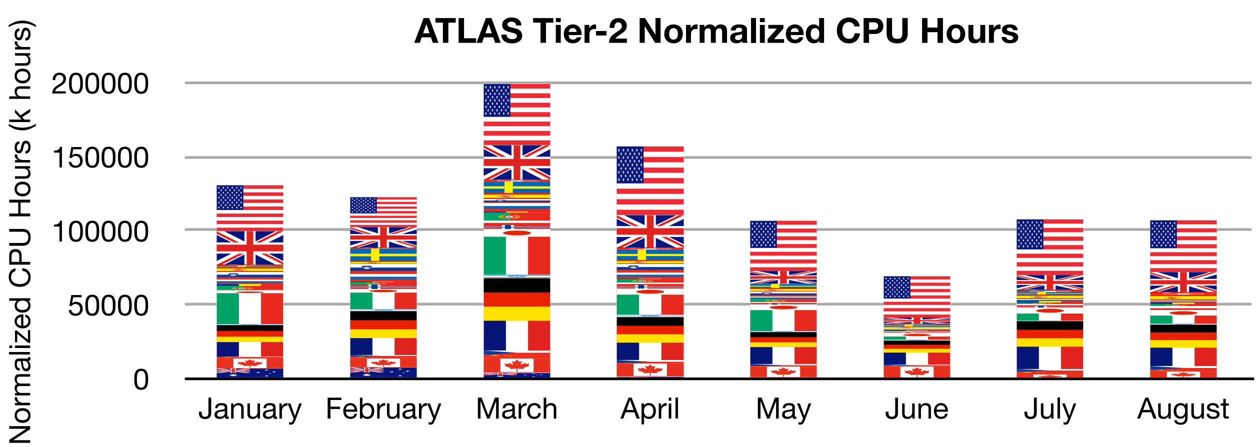


Figure 6: ATLAS CPU Time Used on Tier-2 sites in 2010

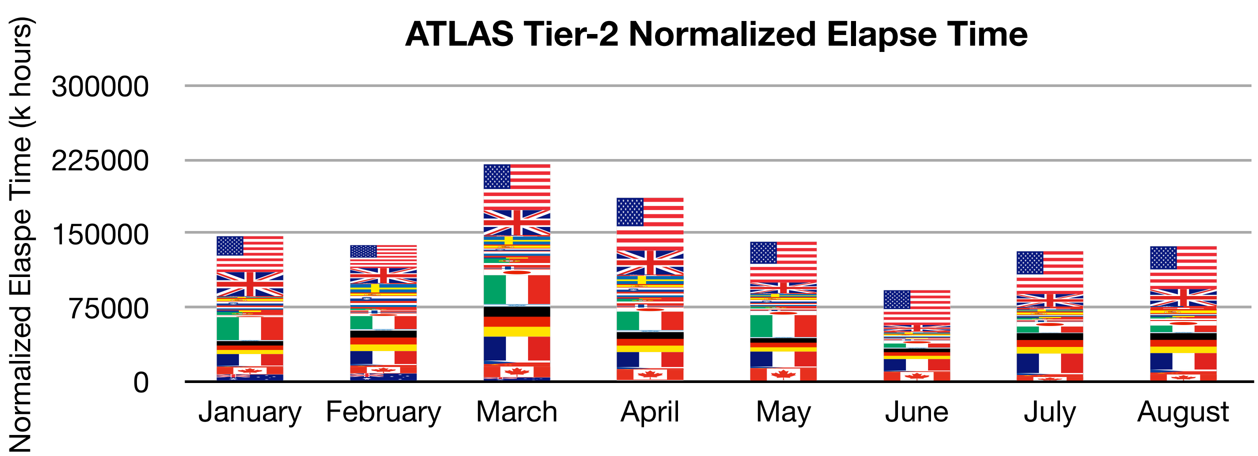


Figure 7: ATLAS Wall Clock Time Used on Tier-2 sites in 2010

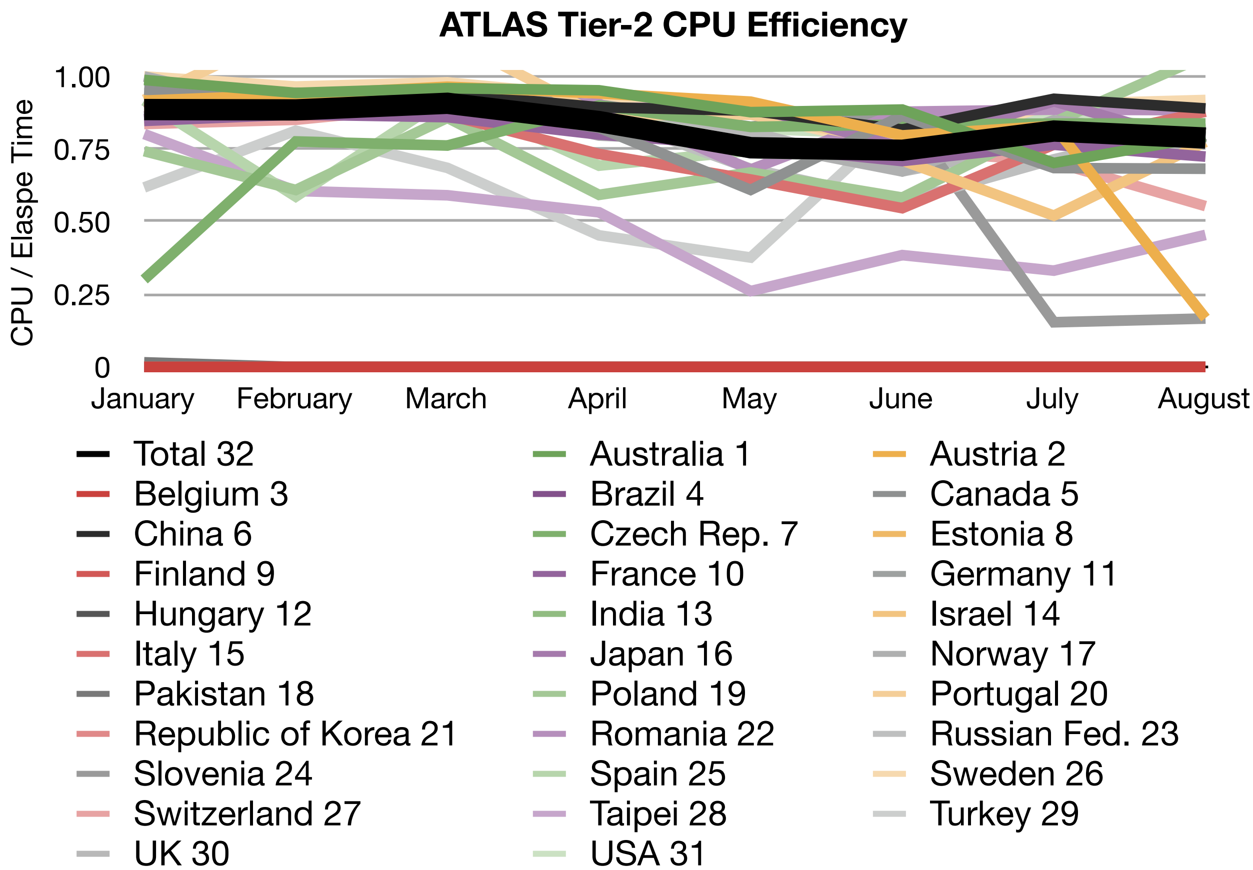


Figure 8: ATLAS CPU Efficiency at Tier-2 Sites in 2010

CMS Reports:

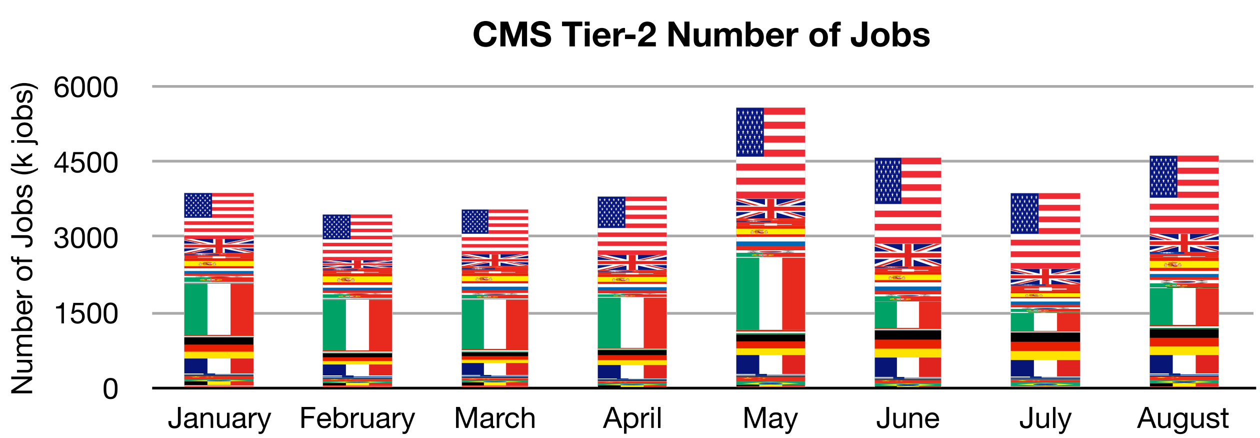


Figure 9: CMS Jobs Executed on Tier-2 sites in 2010

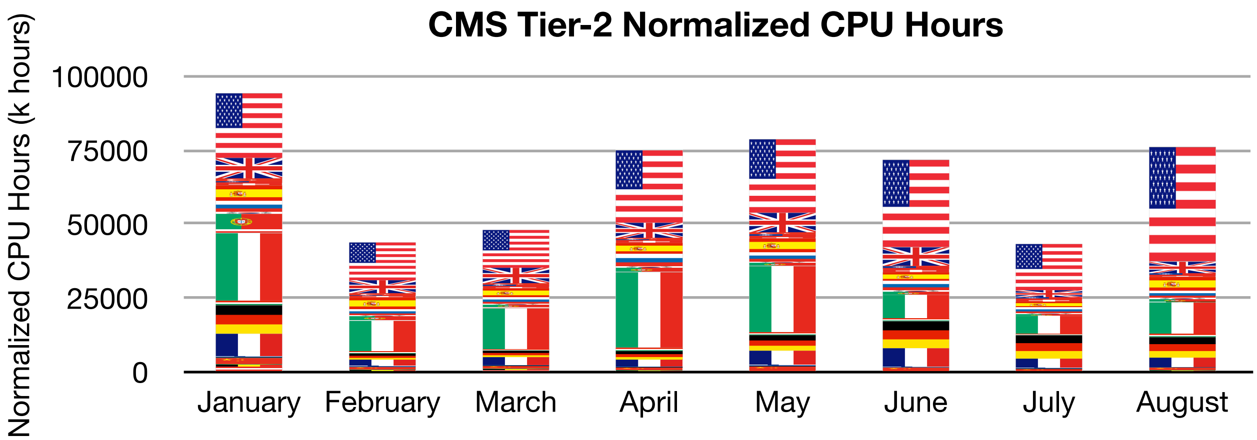


Figure 10: CMS CPU Time Used on Tier-2 sites in 2010

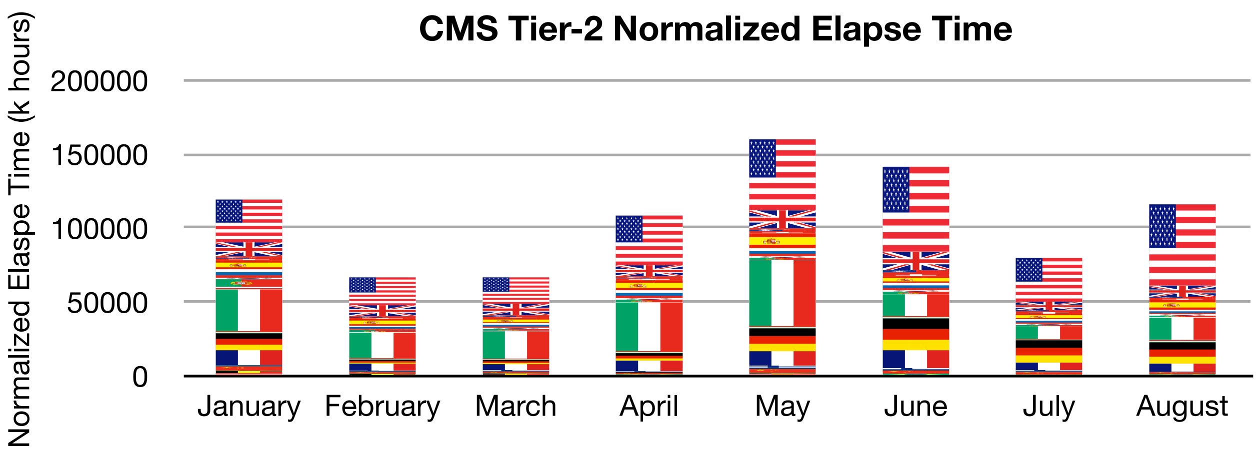


Figure 11: CMS Wall Clock Time Used on Tier-2 sites in 2010

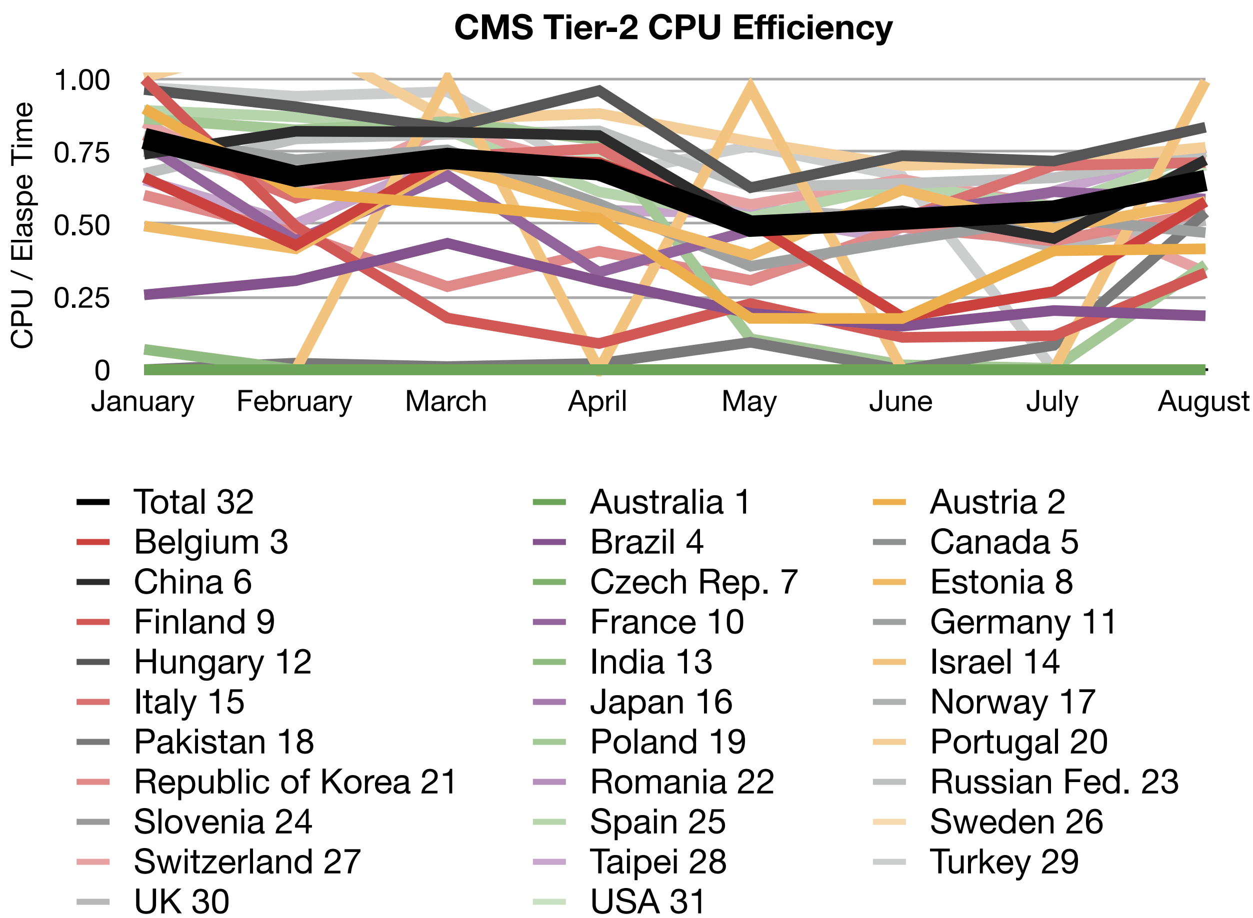


Figure 12: CMS CPU Efficiency at Tier-2 Sites in 2010

LHCb Reports

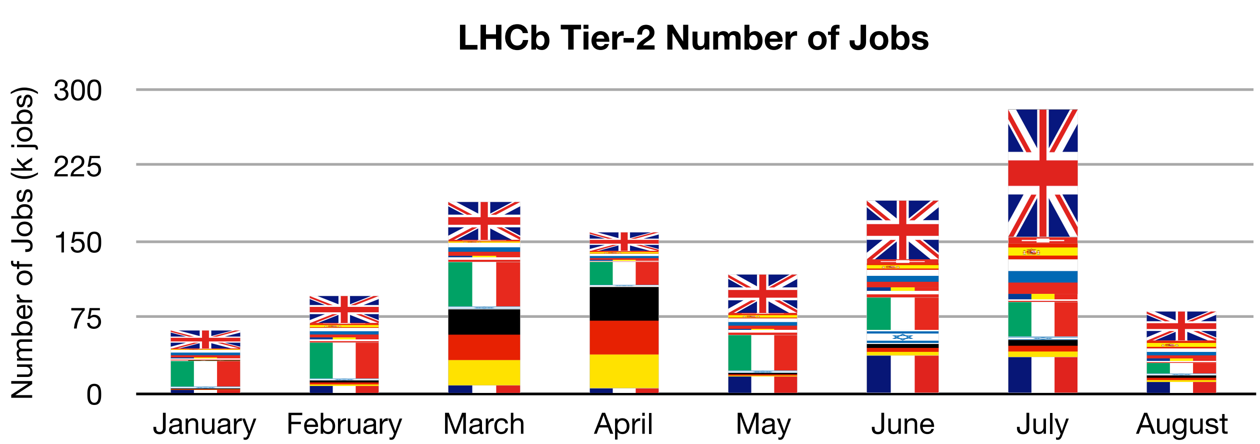


Figure 13: LHCb Jobs Executed on Tier-2 sites in 2010

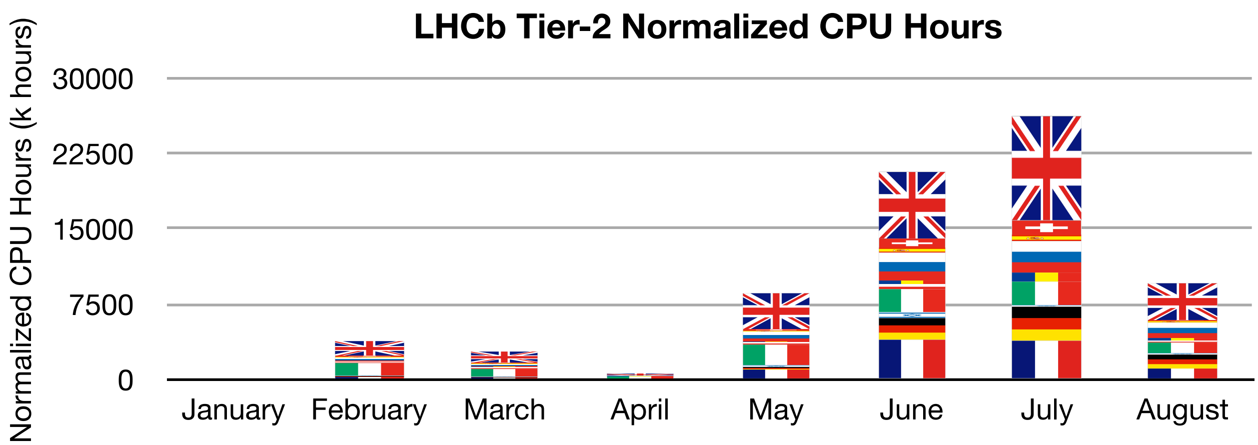


Figure 14: LHCb CPU Time Used on Tier-2 sites in 2010

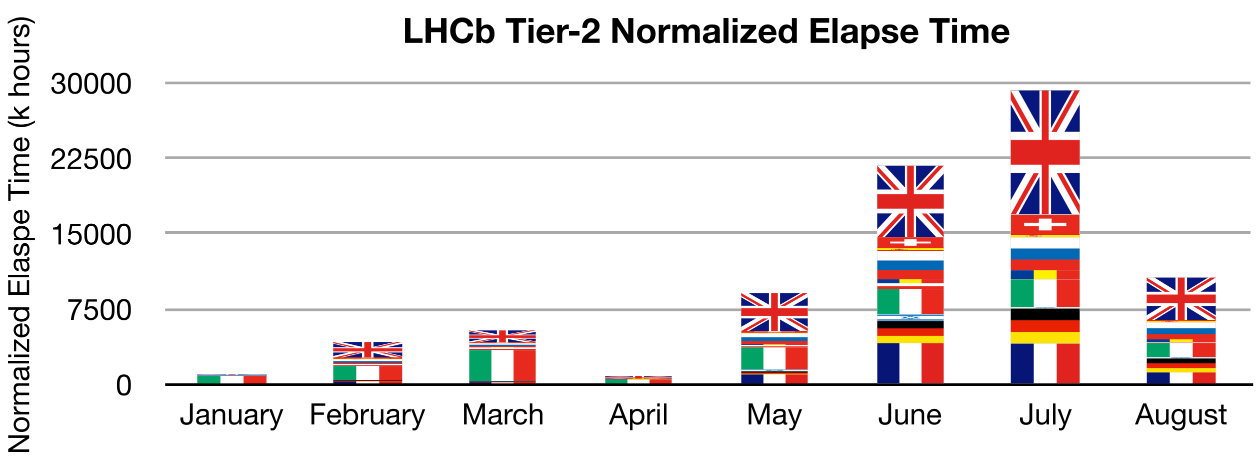


Figure 15: LHCb Wall Clock Time Used on Tier-2 sites in 2010

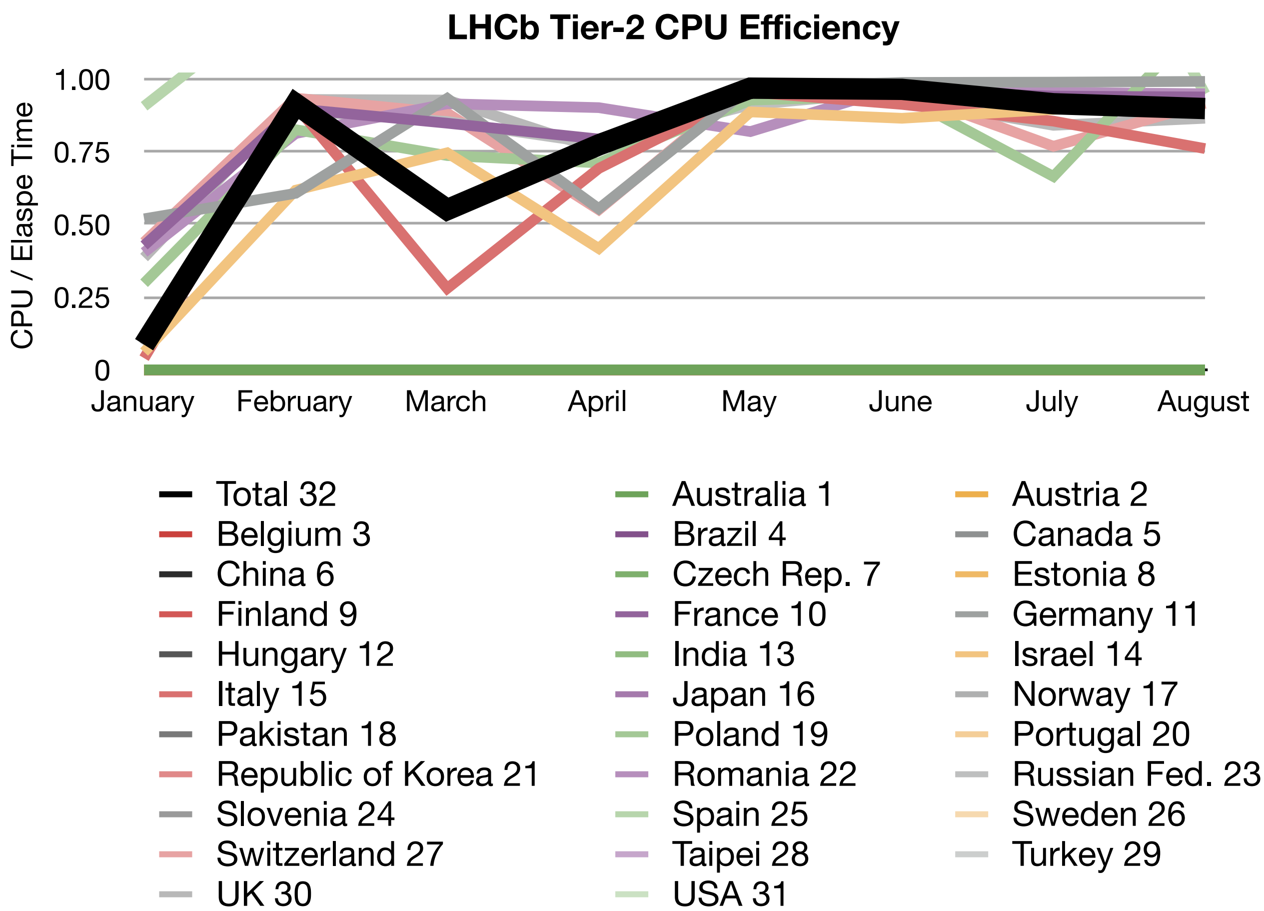


Figure 16: LHCb CPU Efficiency at Tier-2 Sites in 2010

Country Usage Reports for Tier-2s

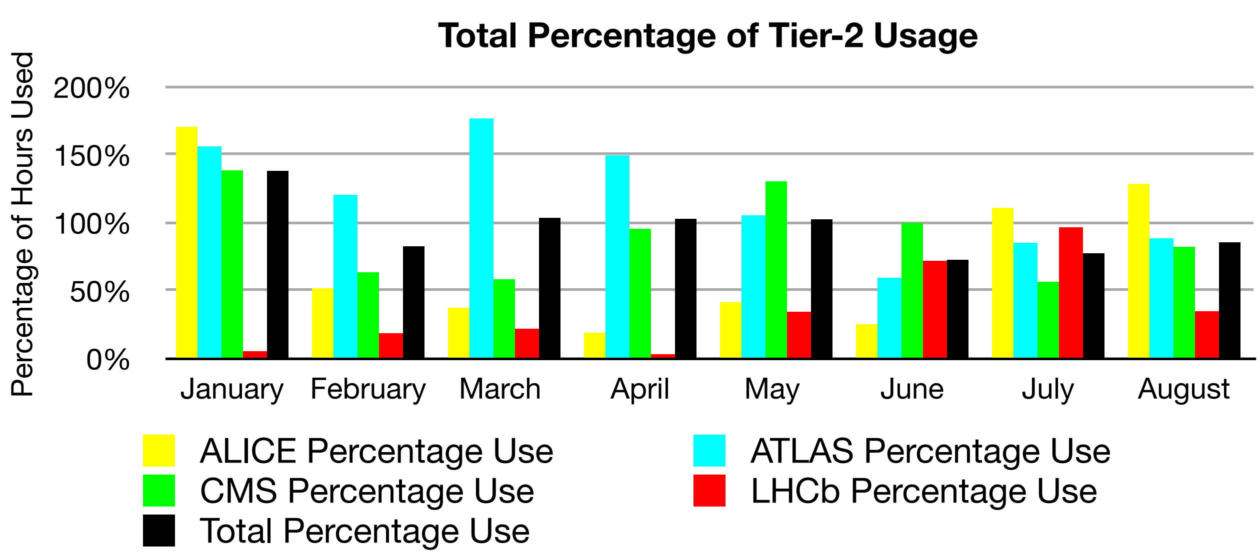


Figure 17: Total Usage of Tier-2 Sites by Experiment

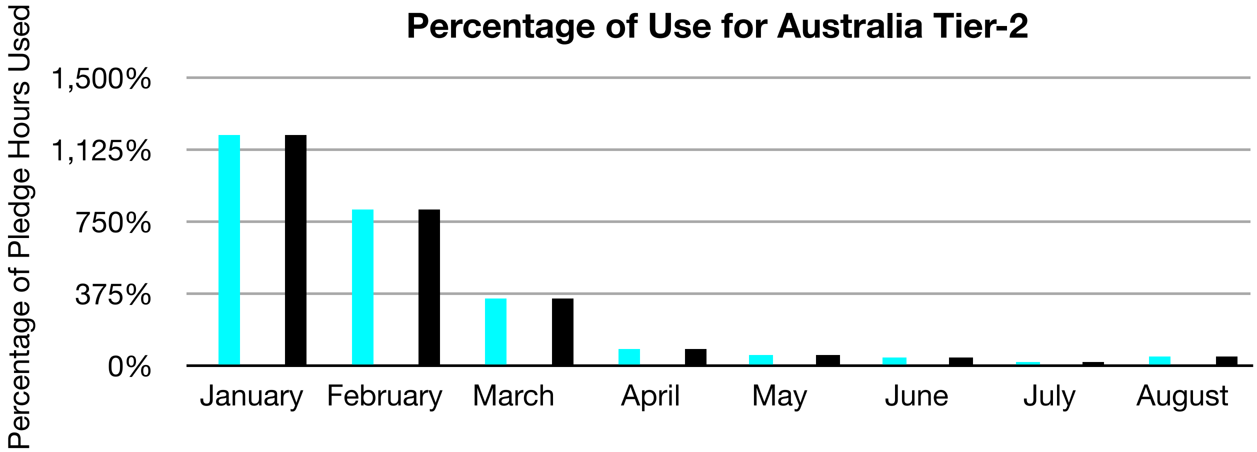


Figure 18: Australia Tier-2 Usage

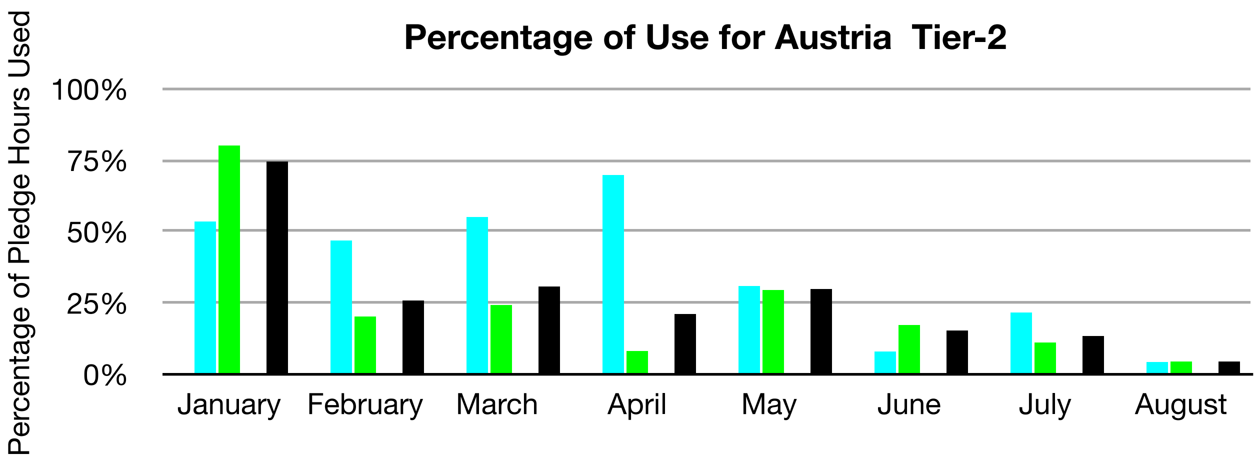


Figure 19: Austria Tier-2 Usage

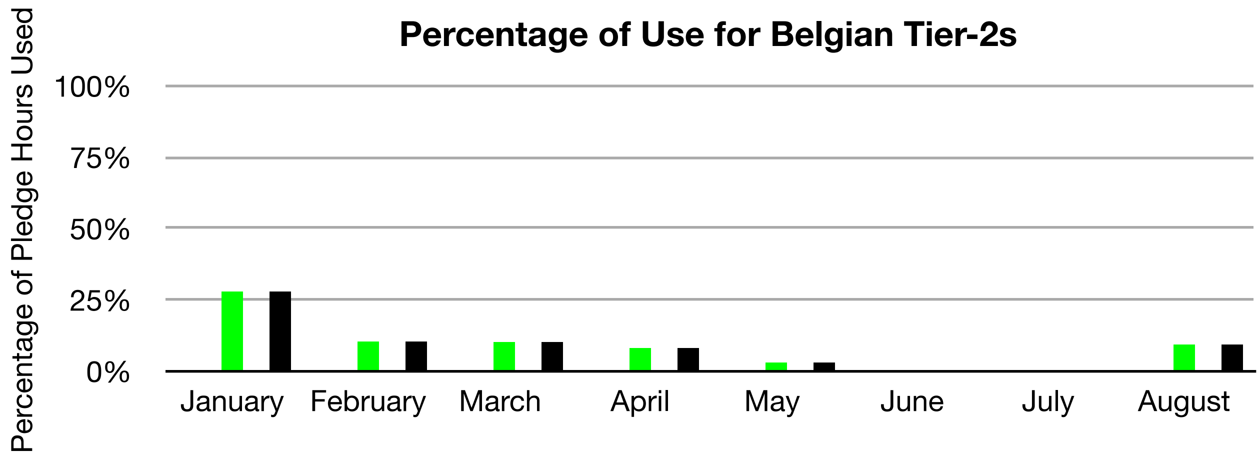


Figure 20: Belgian Tier-2 Usage

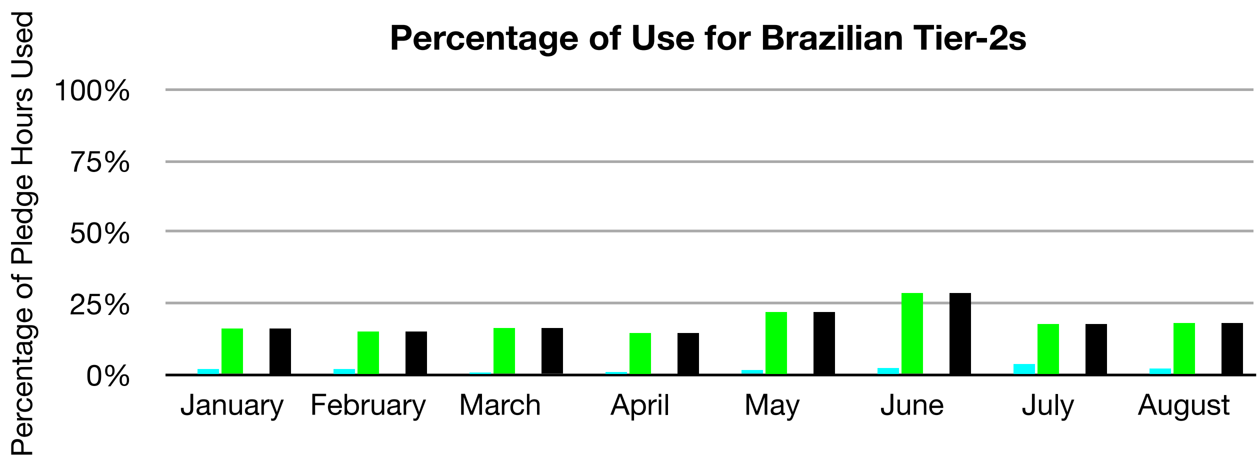


Figure 21: Brazilian Tier-2 Usage

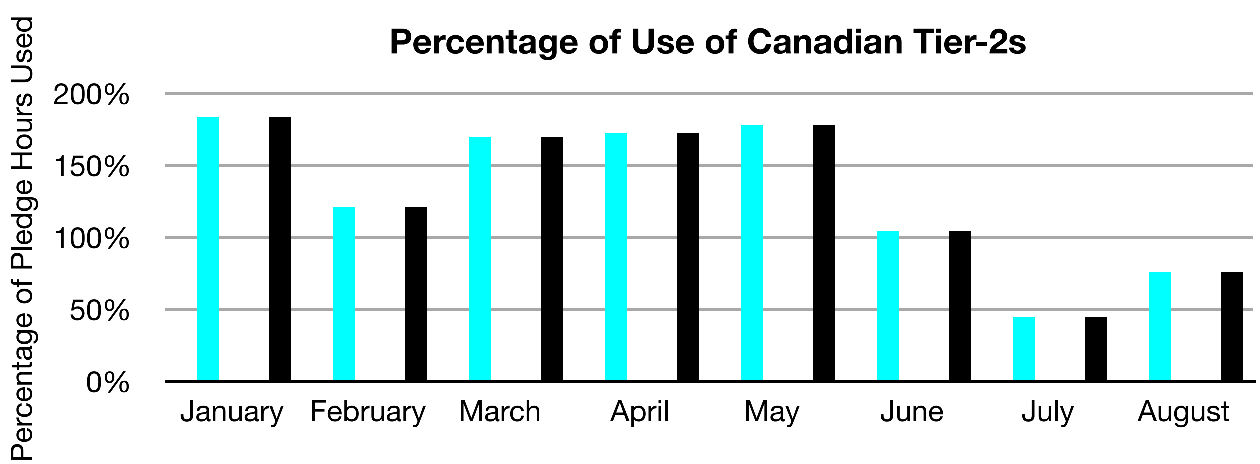


Figure 22: Canadian Tier-2 Usage

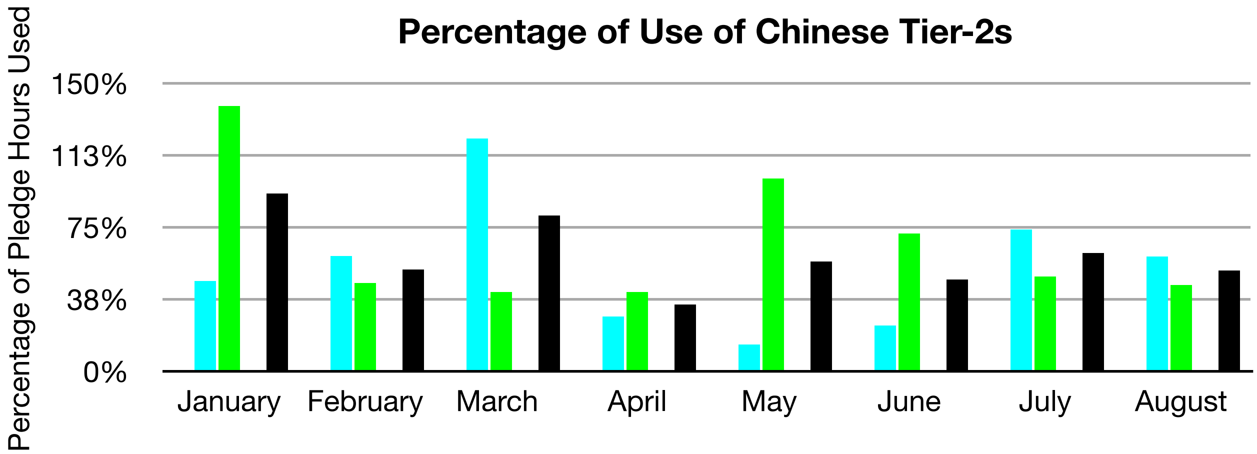


Figure 23: Chinese Tier-2 Usage

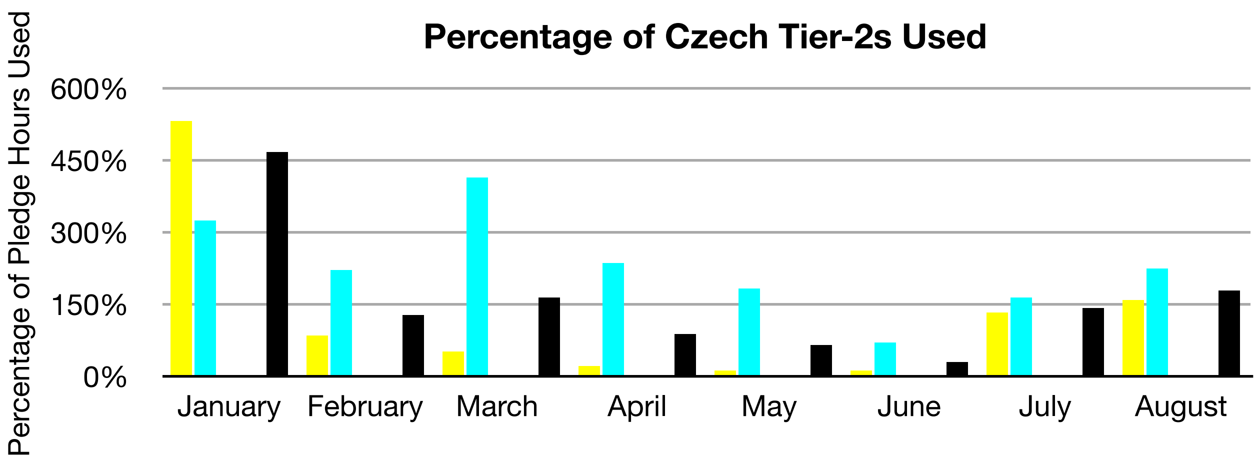


Figure 24: Czech Tier-2 Usage

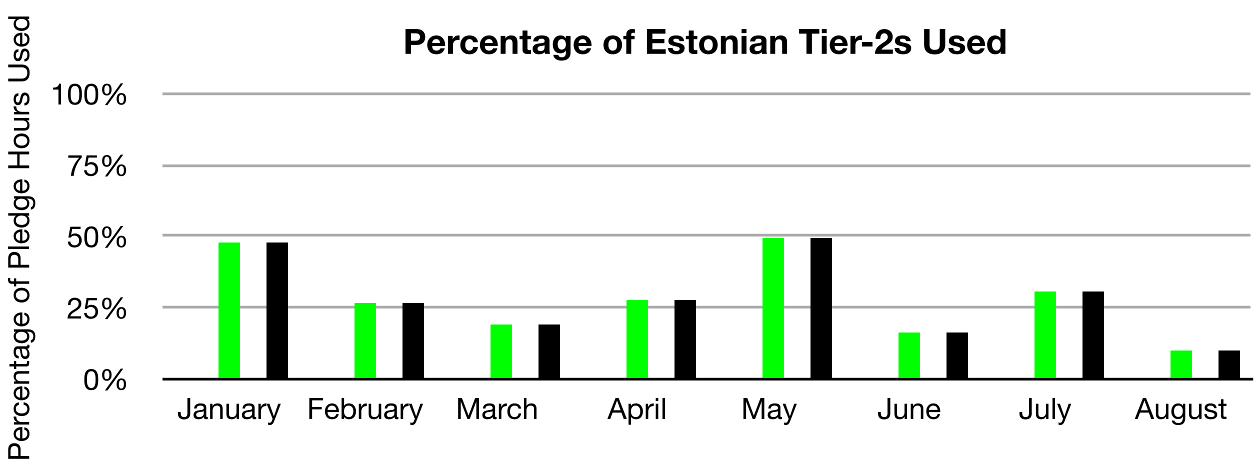


Figure 25: Estonian Tier-2 Usage

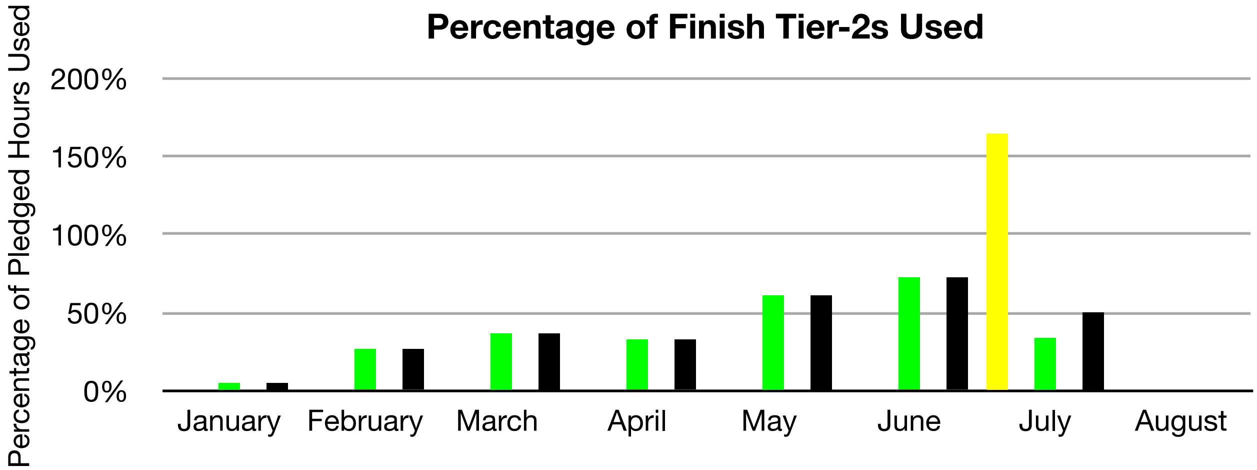


Figure 26: Finnish Tier-2 Usage

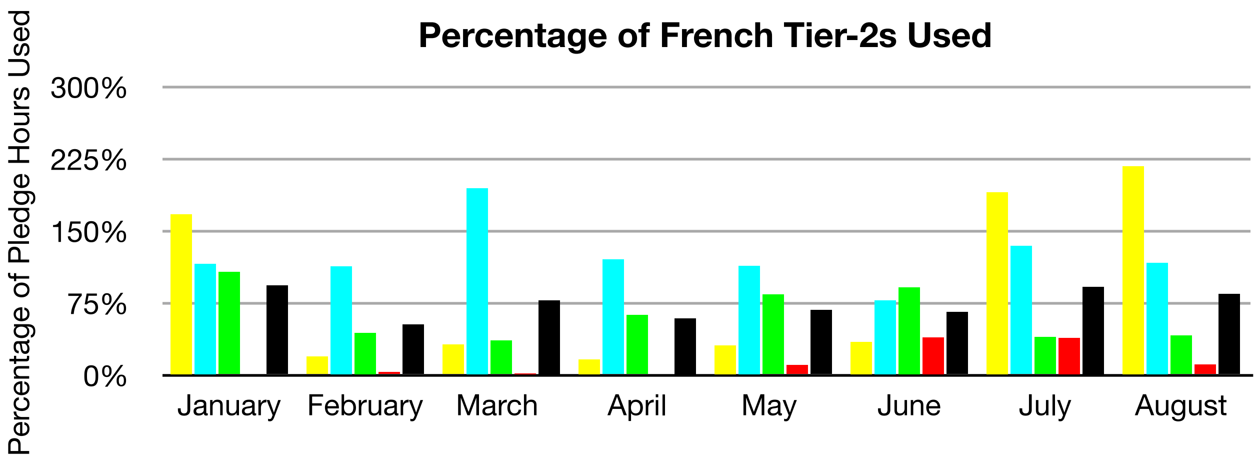


Figure 27: French Tier-2 Usage

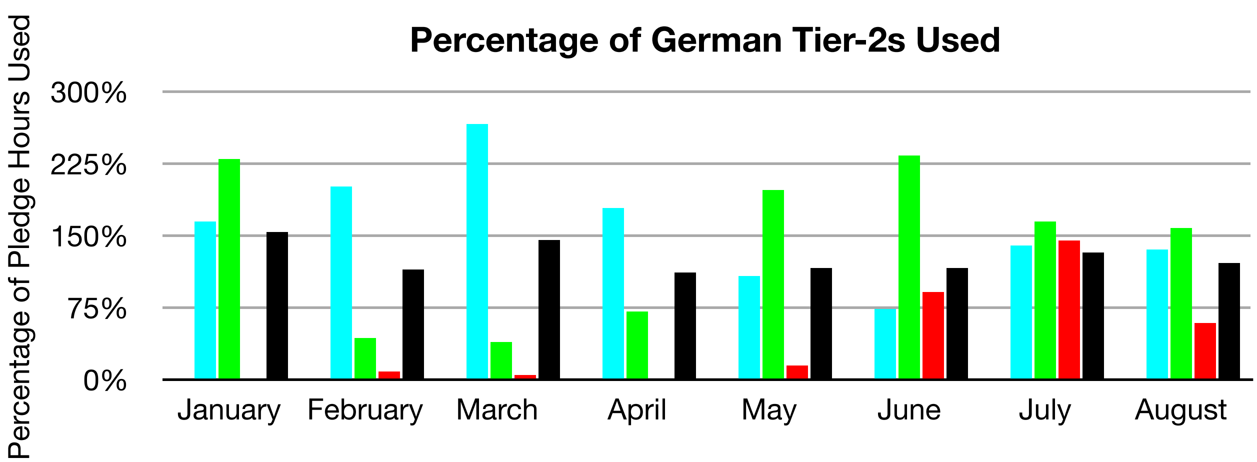


Figure 28: German Tier-2 Usage

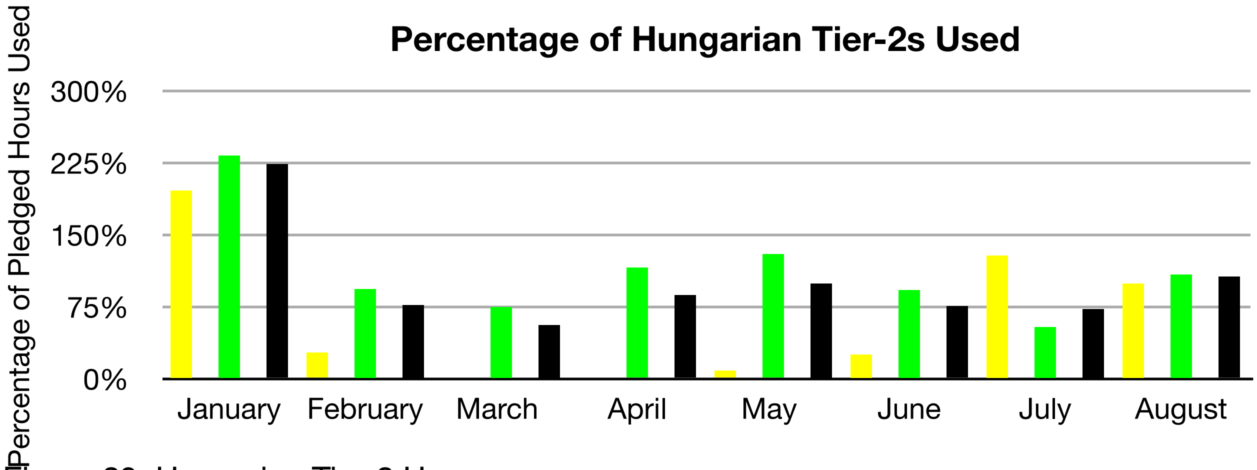


Figure 29: Hungarian Tier-2 Usage

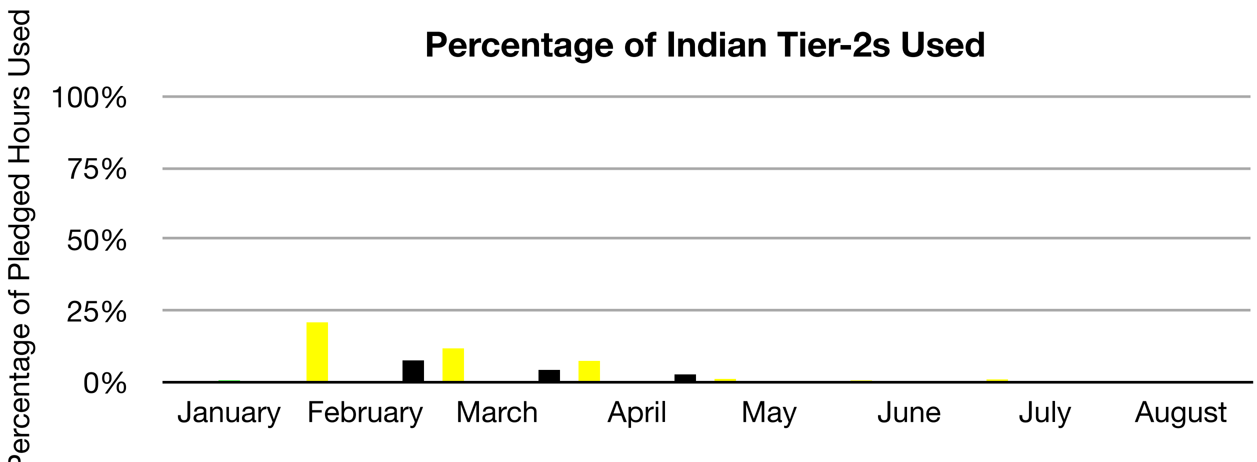


Figure 30: Indian Tier-2 Usage

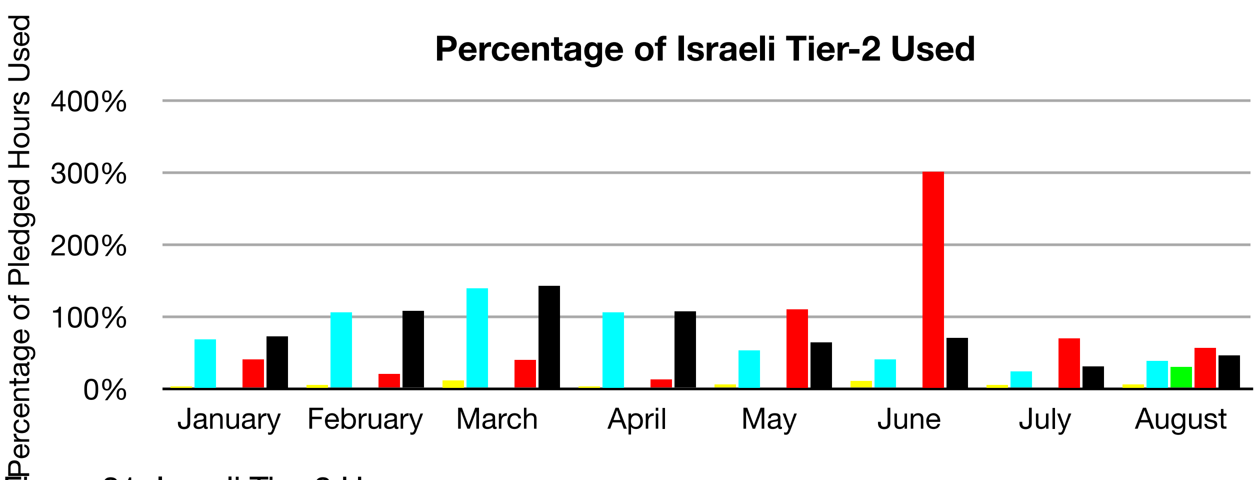


Figure 31: Israeli Tier-2 Usage

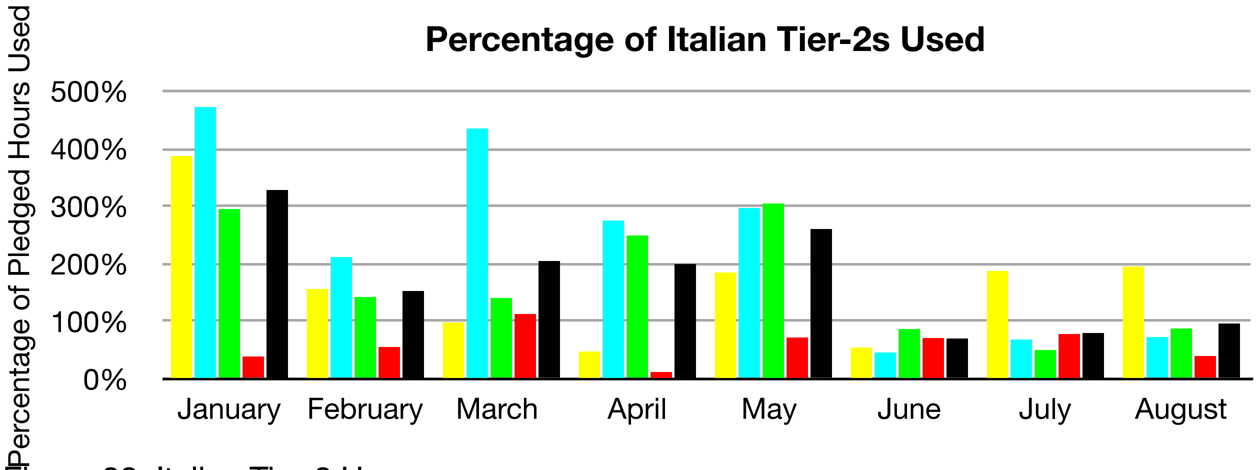


Figure 32: Italian Tier-2 Usage

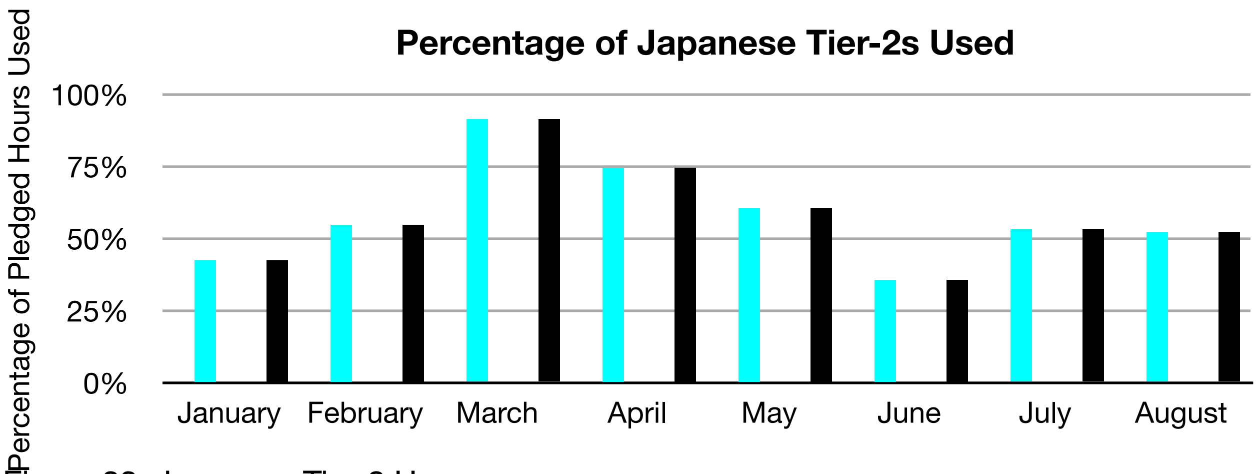


Figure 33: Japanese Tier-2 Usage

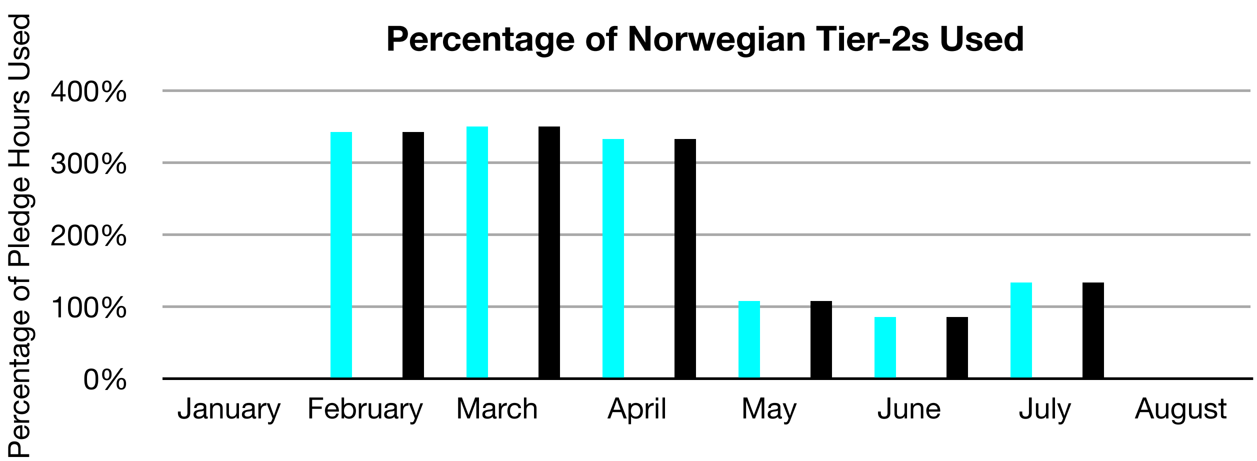


Figure 34: Norwegian Tier-2 Usage

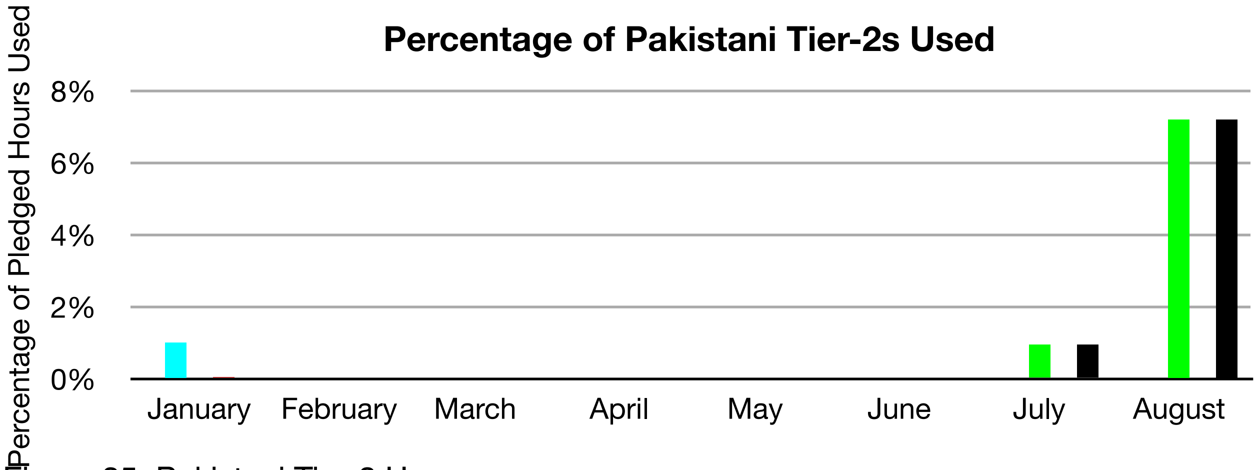


Figure 35: Pakistani Tier-2 Usage

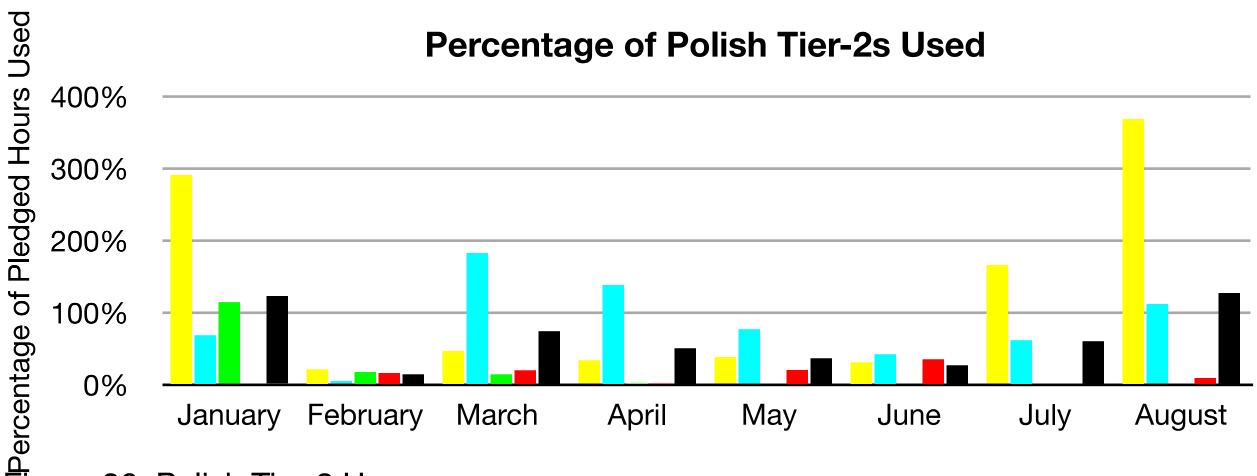


Figure 36: Polish Tier-2 Usage

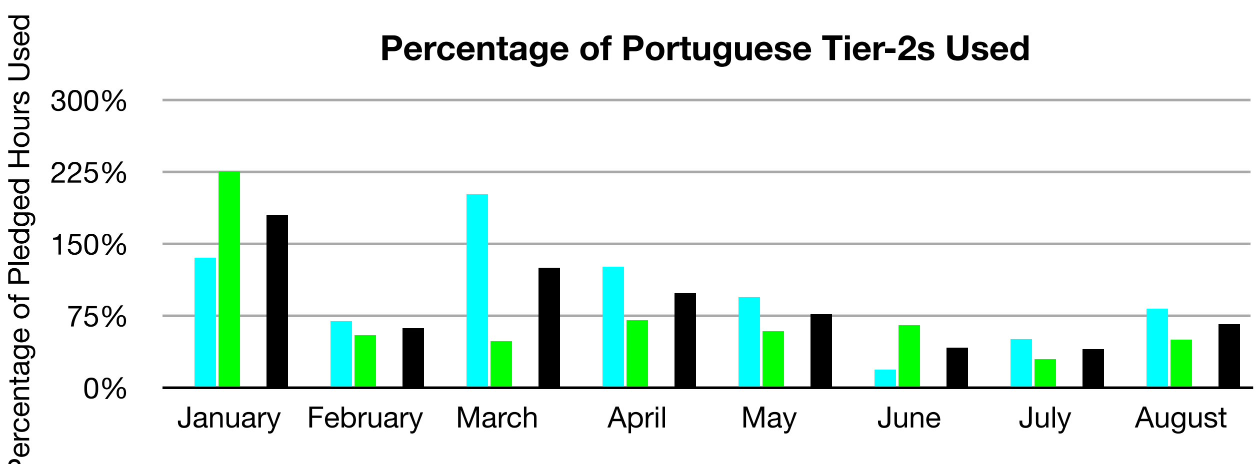


Figure 37: Portuguese Tier-2 Usage

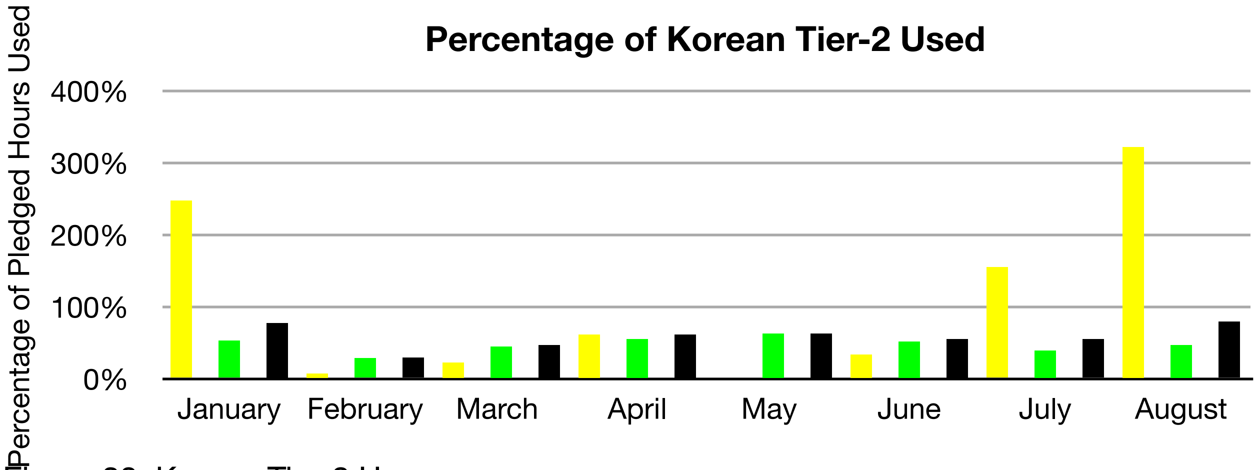


Figure 38: Korean Tier-2 Usage

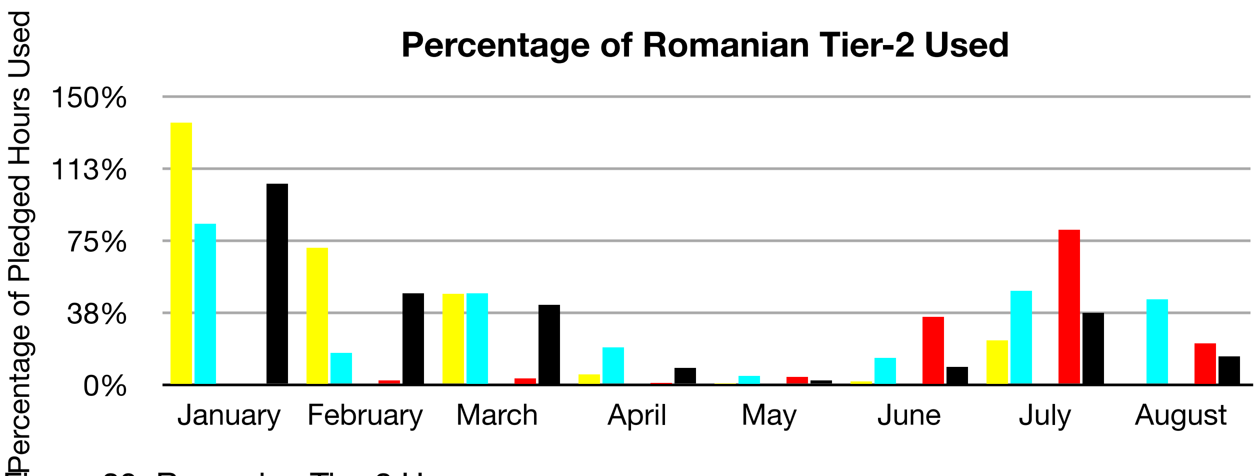


Figure 39: Romanian Tier-2 Usage

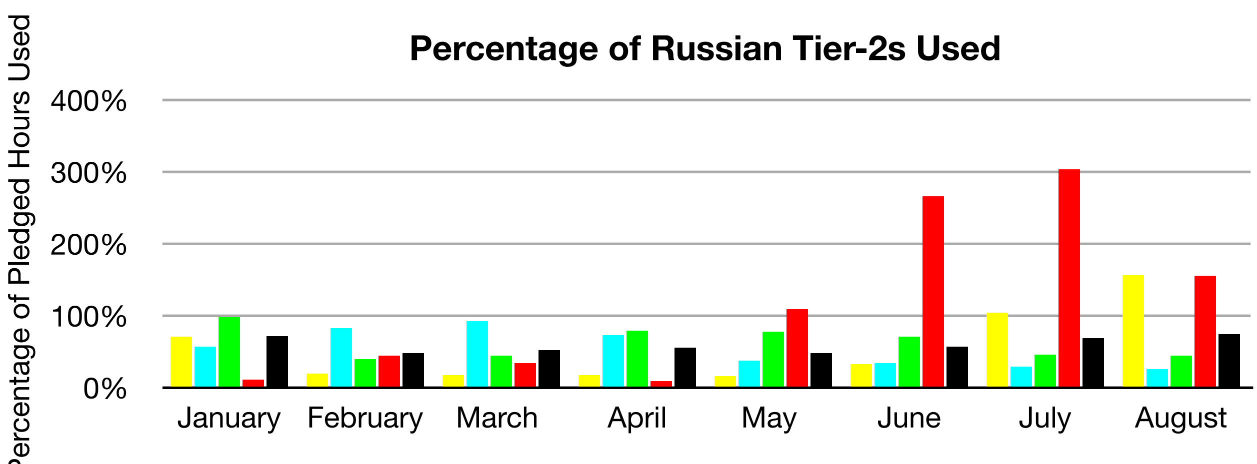


Figure 40: Russian Tier-2 Usage

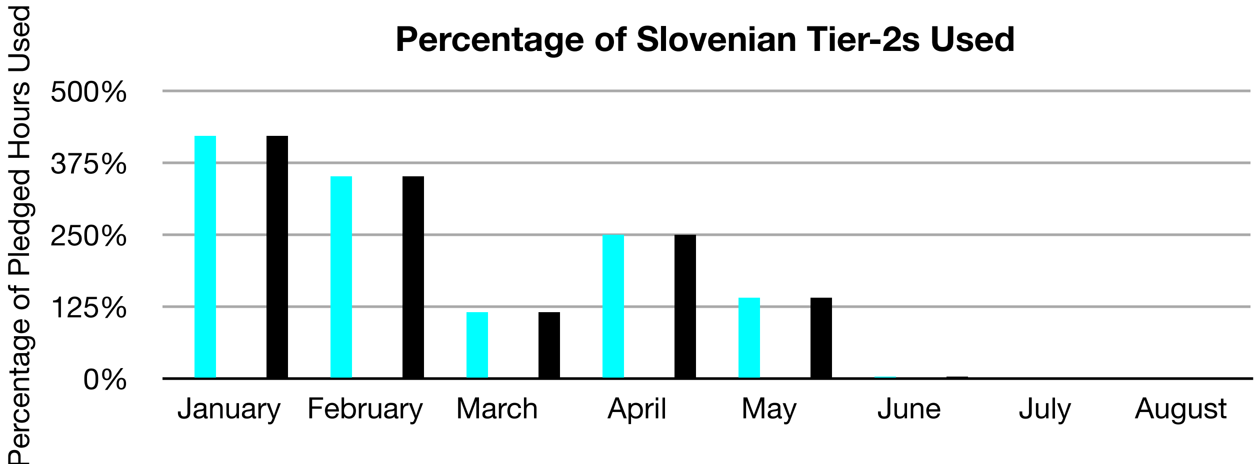


Figure 41: Slovenian Tier-2 Usage

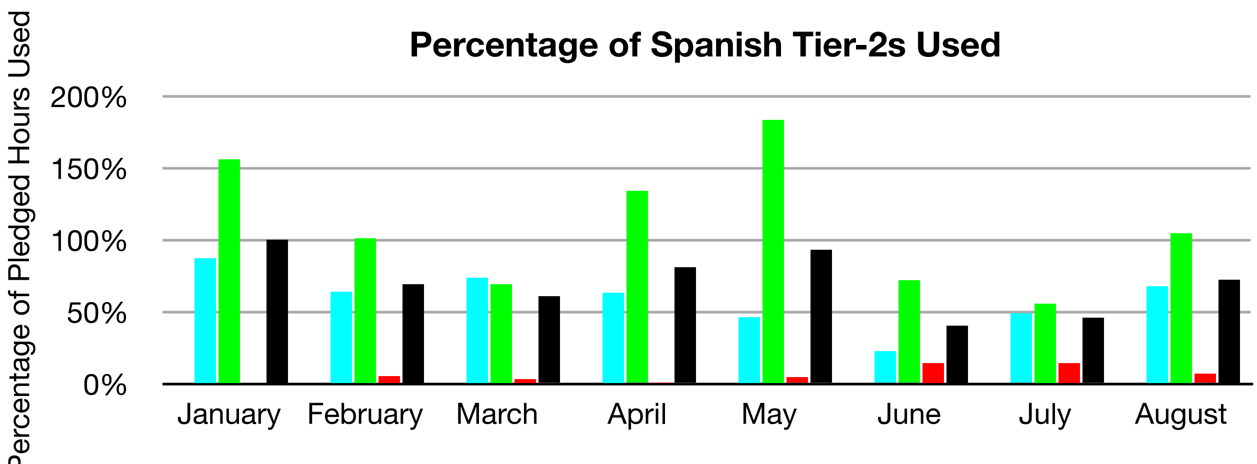


Figure 42: Spanish Tier-2 Usage

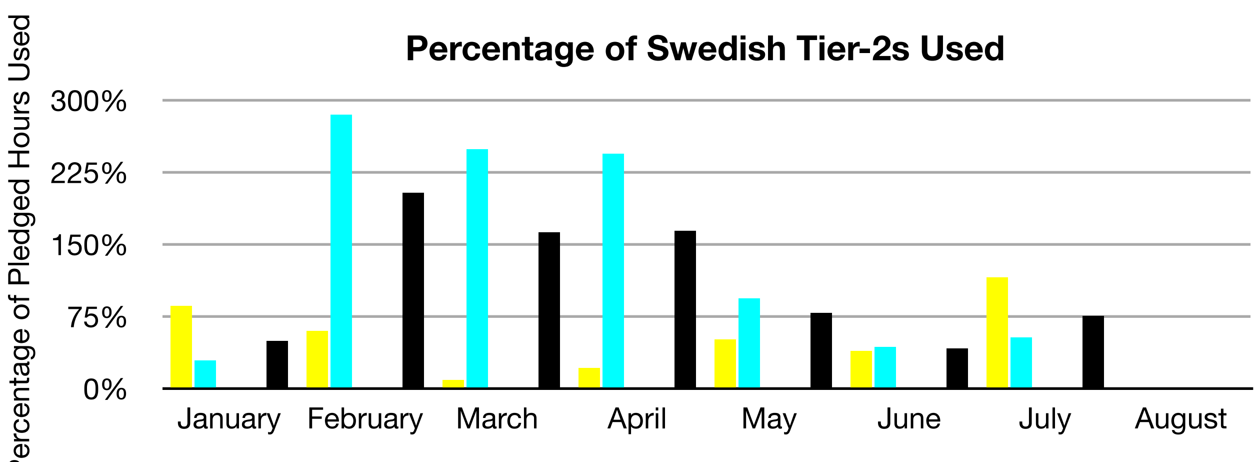


Figure 43: Swedish Tier-2 Usage

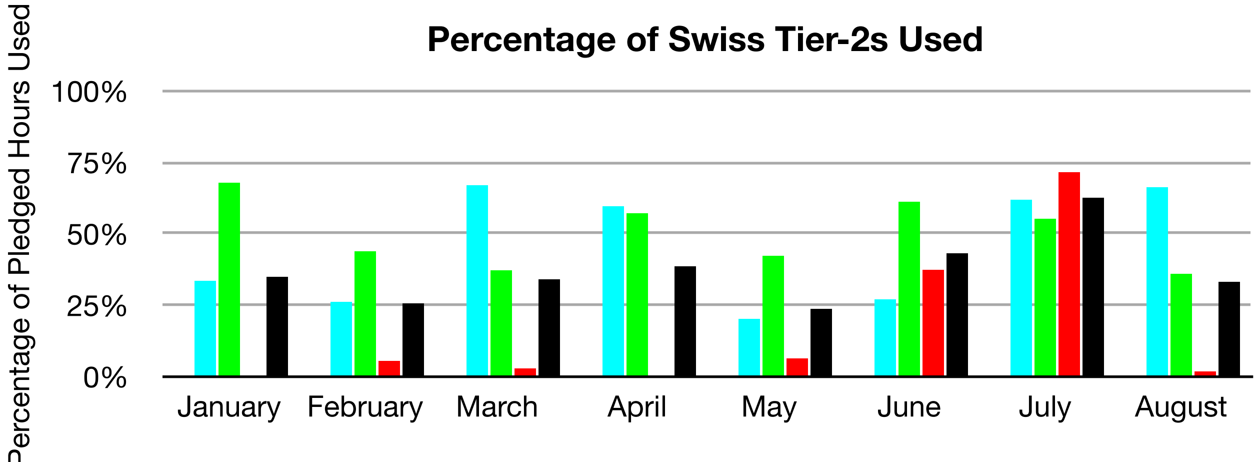


Figure 44: Swiss Tier-2 Usage

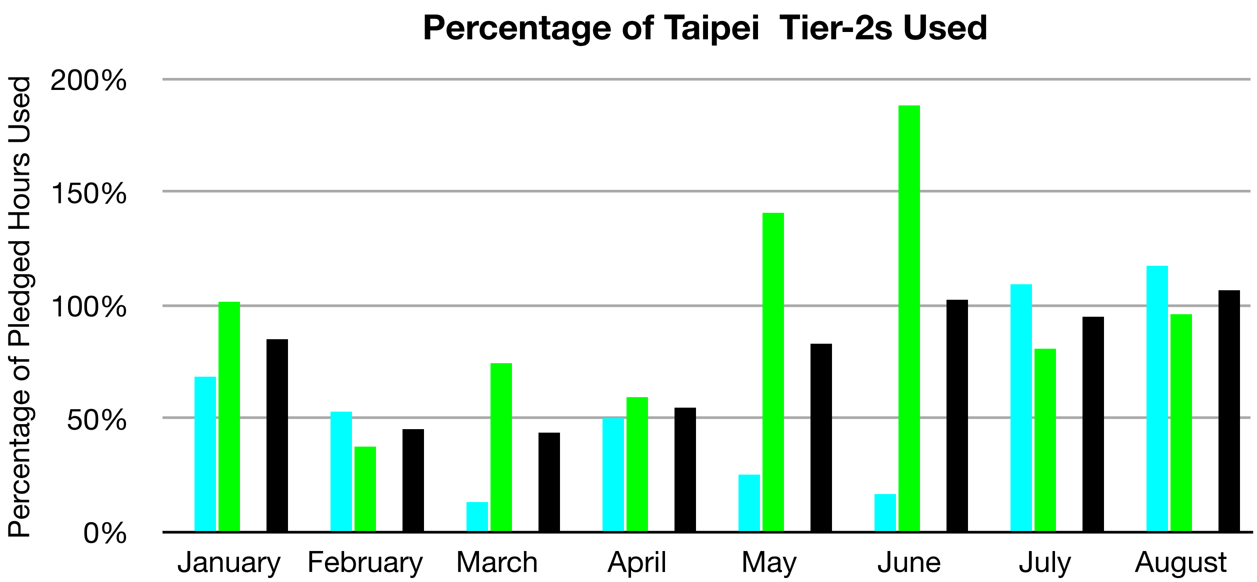


Figure 45: Taipei Tier-2 Usage

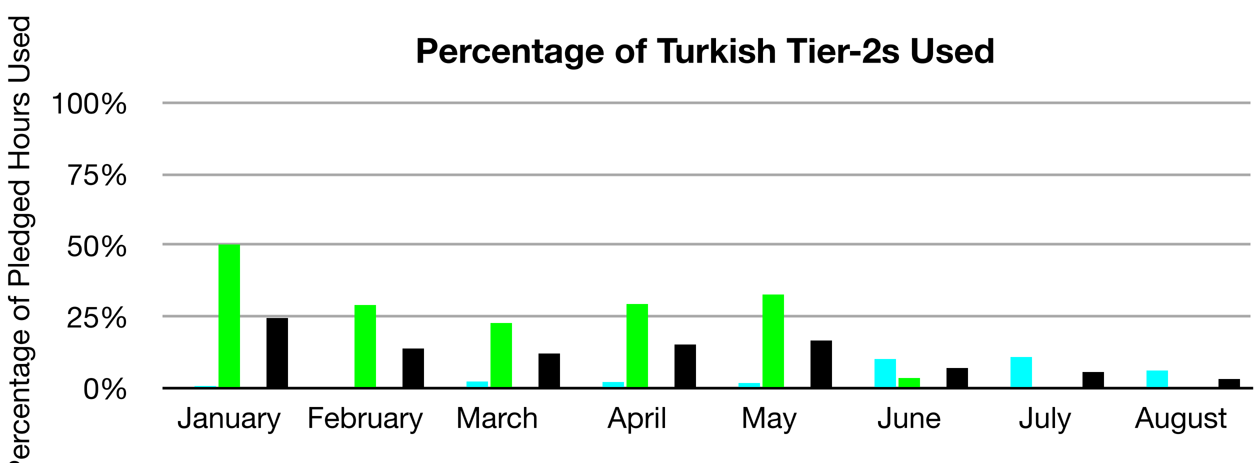


Figure 46: Turkish Tier-2 Usage

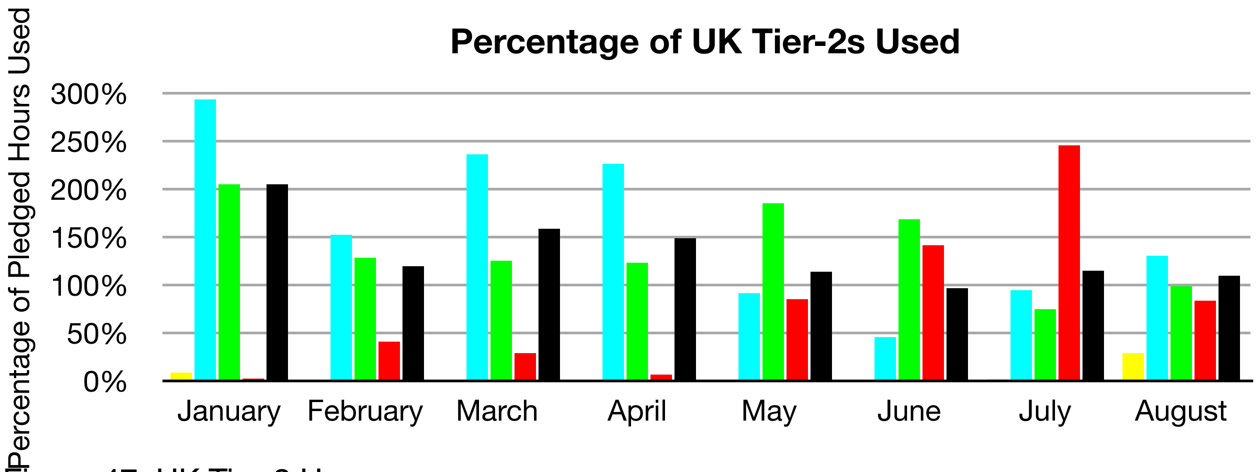


Figure 47: UK Tier-2 Usage

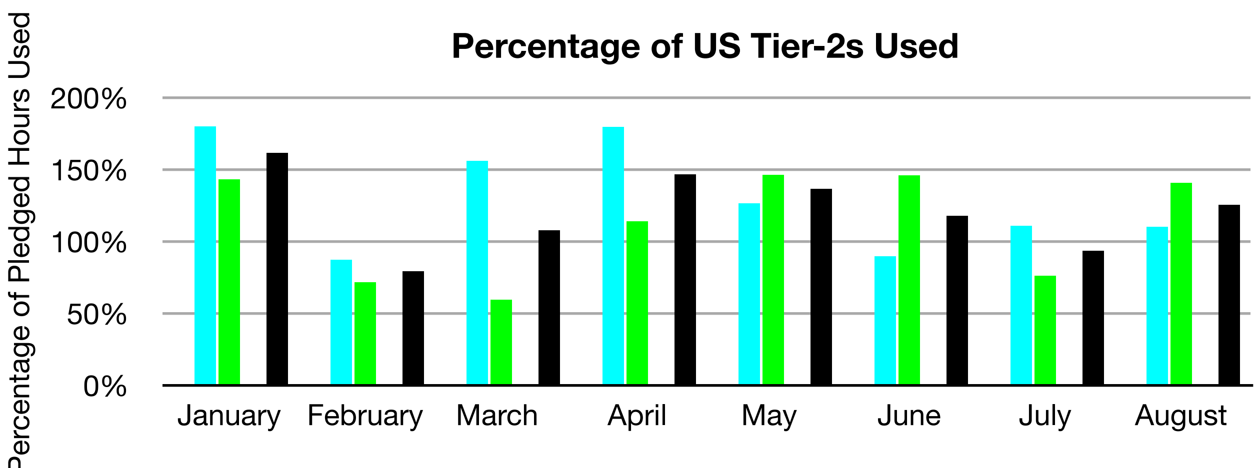


Figure 48: US Tier-2 Usage