Advanced Virgo Computing Needs 2017

This document focuses the CPU, disk and tape needs in Virgo computing centres (CCIN2P3, CNAF and Nikhef) for the Data Analysis of LIGO/Virgo data in 2017. It is based on the information delivered by DA contact persons to the VDAS team.

Status (2016)

2016 has been a year of transition for the usage of computing resources by Virgo collaboration, with different effects on the computing centres accessible by Virgo collaborators. The analysis of the O1 data generated a substantial increase of the usage of CNAF mainly starting from September 2016. The implementation of the accounting system allow to show that CNAF is in the week 8-15 November 2016 the 3rd computing centre of the LVC collaboration in terms of CPU core hours and the 5th since O1 (Figure 1). In fact the CW analysis asked for an increase of the computing power from 25kHS06 (Virgo pledge) to 35kHS06, kindly offered by CNAF clustering the resources unused by other CSN2 experiments. This fully justify the request for 2017 to have 36kHS06 at CNAF, considering the analysis of the O2 data.

CLUSTER PAGE: shows a summary of CPU time used at each site

	Cluster	CPU CORE HOURS (7 days) ▼	CPU CORE HOURS (52 weeks)	CPU CORE HOURS (total)	CPU CORE HOURS (since O1)
0	Total	934,272	148,641,803	222,669,009	172,675,014
1	LIGO-CIT	391,525	20,574,675	30,767,194	24,172,659
2	NEMO-UWM	249,885	28,760,888	37,883,362	33,795,617
3	VIRGO.CNAF	105,004	10,883,093	10,883,093	10,883,093
4	LIGO-LLO	91,316	989,683	3,047,851	1,194,874
5	IUCAA	47,366	152,871	152,871	152,871
6	ARCCA-CDF	36,016	8,220,048	11,572,100	10,160,102
7	LIGO-LHO	13,154	834,898	2,193,446	1,131,320
8	ATLAS-AEI	6	63,486,432	107,889,526	74,901,302
9	SUGAR-SU	0	14,739,215	18,279,566	16,283,176

Figure 1 - Snapshot of the LVC accounting system: https://ldas-gridmon.ligo.caltech.edu/ldg_accounting/cluster_page.html.

In 2016, the Virgo collaboration has requested to CCIN2P3 5M HS06 hours. So far only 1.1 M HS06 hours has been used. The main reason is that only few users are regularly using resources at CCIN2P3 : the EM-follow up Virgo team lead by Barbara Patricelli (Pise), the cosmic string search team lead by Florent Robinet (Orsay) and the long transient search lead by Marie Anne Bizouard (Orsay). The most CPU demanding search, the long transient search, has not run at CCIN2P3 in 2016 because the pipeline has not yet been modified to be able to run there. This explains the discrepancy between the demand and the usage.

All experimental data (LIGO and Virgo) are in HPSS (in total more than 900 TB of data). A 15 TB disk space is available for the jobs' outputs. 200 GB of disk in the $GROUP_DIR disk are is used by Virgo for providing a web access to Virgo search outputs.

The CCIN2P3 is now able to execute pyCBC grid jobs successfully, but only few jobs have been submitted so far, although the information has been circulated.

A new entry in 2016 has been the use of the Nikhef computing center. In 2016 the Rome CW pipeline has been adapted to run via EGI at Nikhef and up to 1000-1500 jobs have been launched in parallel in some time periods (Figure 2).
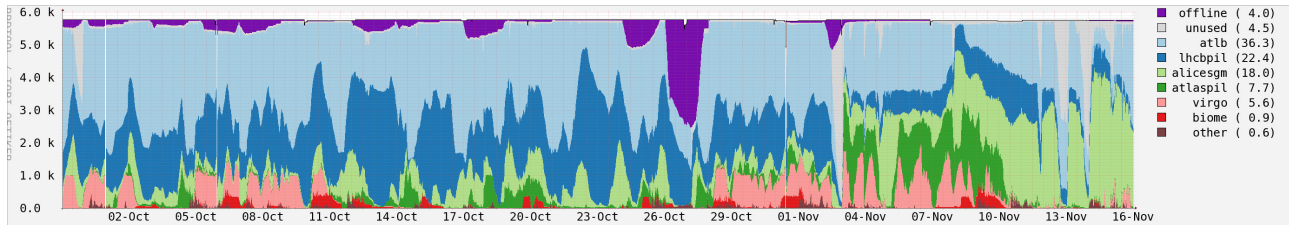

Figure 2 - Usage of the Nikhef computing centre in the last month. In pink the Virgo jobs.

Needs 2017

In 2017 the main user of the Virgo computing centre will be again the CW analysis but new entries are expected. cWB is attempting to move part of the analysis at CNAF, pyCBC is capable to access CCIN2P3 and CNAF via OSG (soon also cWB will be able to do it) and the long transient search will run at CCIN2P3. The resources declared by each pipeline are described in Table 1.

Table 1 - Resources requested by each pipeline

Pipeline	CPU Needs (power-cores)	Disk (TB)	CCIN2P3		CNAF		EGO		Nikhef	
			CPU hours (energy)	Disk (TB)	Cores (power)	Disk (TB)	Cores	Disk	Cores (Power)	Disk (TB)
CW All sky	3500	65			2500	50			1000	15
cWB	145	1			145	1				
Stampas	11	5	1000	5						
Cosmic String	5		440							
EM-follow up	5		440							
Glitches /detchar	10	2					10	2		
Online data quality	16						16			
Omicron	16						16			
Monitorweb/ VIM	32						2x16			
Monitorweb/ VIM	16						16			
Silentec	32						2x16			

Data Transfer and Storage

The amount of data generated by O2 is surely out of specification, considering that both LIGO and Virgo are writing more than described in the specs. Let make here a simple evaluation:

Virgo is writing currently 36MB/s in commissioning mode. Let expect that in science mode we will write 30MB/s. If O2 duration for Virgo is 4 months Virgo will write 312TB (bulk data) and an additional 10% of trend and reduced data, for a total of about 350TB. LIGO is writing 26MB/s for each interferometer instead of the 11MB/s of O1. The O1 data currently transferred from LIGO occupies 1.5TB, hence considering that the run will be two times longer and

accounting the increased data rate will be safe to account for about 20 TB from LIGO (considering also multiple calibrations). The need of science data storage for Virgo in 2017 is 370TB.

Currently at CNAF there are 590TB of pledge with an occupation of 66% (388TB); it seems reasonable move on tape all the old files allowing the accommodation of the additional 370TB with only additional 64TB.

At CCIN2P3 experiments data are stored on tapes and retrieved via xrootd thanks to a disk cache system, which is shared with other experiments. Usually the size of the cache disk used by the experiment is roughly 30% of the total volume of accessed data on tapes. Assuming a volume of 370 TB accessed in 2017, the xrootd cache disk used by Virgo will amount to ~100 TB. Jobs' outputs are stored in the semi-permanent disk space /sps/virgo/. We have currently 15TB of disk space and a small increment of 5TB is requested in 2017.

At EGO we had in 2016 a crisis on the availability of disk space in the "Procdata" volume; the problem has been mitigated implementing a quota control and monitor; but further needs, related to the next scientific run(s) are expected and it is mandatory to foresee an increase of 100TB in the storage in Cascina.

In the following tables are listed the additional requests at the computing centres:

Table 2 - 2017 needs at CNAF

CPU power [HS06]	Disk Storage [TBN]	Tape Storage (additional TB)
36000 (pledge)	656 (pledge)	300 (To be verified)

Table 3 - 2017 needs at CCIN2P3

CPU Energy [HS06.hours]	Disk Storage (TBN)	Tape Storage (additional TB)
2 M HS06.hours	5TB additional to the 15TB currently available	370

Table 4 - 2017 needs at Nikhef

CPU Energy [HS06]	Disk Storage (TBN)	Tape Storage (TB)
25000	20	0