CNRS

Centre National de la Recherche Scientifique

INFN

Istituto Nazionale di Fisica Nucleare

Virgo Computing needs for 2010

VIR-0640A-09

Release: 5

Date: 22 October , 2009

VIRGO * A joint CNRS-INFN Project
Via E.Amaldi, I-56021 Santo Stefano a Macerata, Cascina (Pisa), Italy
Secretariat: Telephone (+39) 050 752 521 * FAX (+39) 050 752 550 * e-mail: W3@virgo.infn.it

CHANGE RECORD

Issue/Rev	Date	Section affected	Reason/remarks
1	06/10/2009		Initial release by GM Guidi
2	08/10/2009	Data transfer	Antonella Bozzi
3	13/10/2009		GM Guidi
4	22/10/2009	2, 5.1, 5.2.2	GM Guidi

1 Introduction

This note describes in short the computing resources and the required computing needs for 2010.

2 Overall computing strategy

The purpose of each computing site involved in the process of the Virgo data analysis is recalled in the Virgo notes VIR-016A-08, VIR-088A-08.
While the two national Computing Centers (CC) of CNAF/Bologna and CCIN2P3/Lyon are mainly used for off-line analysis, the EGO/Cascina site is the data production place in which the data are provisionally stored before being transferred to the final repositories of Lyon and Bologna. Nonetheless, data analysis is also performed in Cascina to support detector commissioning, to monitor the status of the machine and to implement and perform the on-line/in-time analysis for all transient signals.

Most of the analyses are being carried at Bologna and Lyon (which can also be accessed via GRID), but also other computing resources are used. In particular, the use of GRID allows to transparently access other resources both completely or partially dedicated to Virgo. Among these we mention the Roma 416 cores farm, which is the official Virgo Tier-2 site, Pisa, Perugia, APC in Paris, Nikhef and RMKI in Budapest.
At the moment, GRID resources are mainly used for continuous signal and stochastic background searches.

Data are stored in the CCs were they are analysed. In 2009 we had, until about the mid of the year, VSR1 data stored in disk. At the start of VSR2, VSR1 raw data have been stored only in the mass storage systems, and VSR2 data on disk. For the next year 2010, we plan to have all VSR2 data stored on disk. We will thus have about 2 years of data stored on disk for fast access, as already stated (VIR-LIS-ROM-7000-130; EGO-NOT-COM-21).

3 Data Production

3.1 Real Virgo Data

During 2009, Virgo was under commissioninfg until July 14th, when the Virgo Science Run 2 (VSR2) started.
During the commissioning phase, short commissioning runs (February 13-15; March 6-9; June 12-15) have been undertaken in order to check the progress in the commissioning activity. In the computation of the amount of data produced and stored during 2009 we have then to take into account, a part from the VSR2, also the data produced during the commissioning runs and in general the data acquired in Science Mode (SM), as these could be used for Data Analysis studies in framework of the Astrowatch program.
We archive raw data, acquired trend data, 50Hz data and Reduced Data Sets (RDS), which currently amount primarly to h-Reconstructed data. The raw data are currently acquired at a compressed rate of about 11 Mbyte/s, which is larger than the foreseen rate of 8 Mbyte/s considered in VIR-088A-08. Taking into account the Astrowatch data, we have until now (October 1st) a production of new Virgo data in 2009 of about 83 TB. For the last months of 2009 we can foresee a futher increase of about 90 TB [1].

For 2010, owing to the scheduled stop for commissioning of 4 months, we should store about 237 TB (8 months of raw + processed data) + 31 TB (commissioning runs and Astrowatch data) of Virgo data. We foresee to store Astrowatch data directely in the mass storage system.

3.2 LSC data

We have received and stored in the CCs, until now, about 1.4 TB of LIGO data. We can estimate the exchange of 1.9T of data for the next months, for a total of about 3.5T of LIGO data in 2009.

For 2010, supposing a full year of data taking of the two LIGO interferometers, we estimate an exchange of about 8 TB.

[1] **NOTA BENE**: we use powers of 1024 not 1000 to measure sizes.

4 Data transfer

In general, the Virgo and LSC data need to be transferred to the computing centers in a timely and reliable way, so as to enable the users to begin the off-line analysis work as soon as possible.

Virgo data (raw, trend, 50Hz and RDS data) need to be transferred to the CCs in a quasi-continuous way, meaning with a delay of 1-2 days. The collaboration is analysing different alternatives for the tools to be used for the transfer. Tools like bbftpPro and SRB, which were already used during the WSR, are to be used like the engines for the data replication software developed for the data transfers in the 2009. For 2010, the use of other GRID tools like *lcg-utils* is considered, but the collaboration has not arrived to a final decision. The data are transferred simultaneously from Cascina to the CCs according a star architecture owing to capacity of the EGO geographical link.

To transfer LIGO data to the Cascina site and Virgo h_rec data to a LIGO site we use the Lightweight Data Replicator (LDR) of the LIGO collaboration. LIGO data need then to be transferred to the CCs with the same priority of Virgo data.

The tranfer of LIGO h-reconstructed data to Cascina site and of Virgo h-reconstructed data to LSC sites for network on-line analysis is performed also with a different tool (Cm) with a latency of some minutes.

5 Computing and storage in 2009 and 2010

We describe in the following the computing and storage use in 2009 and give an estimation for 2010.

5.1 Computing

The computing needs for 2009, outlined in the VIR-088A-08, were based on the actual use of computing resources during 2008 and on the basis of the possible planned searches on S5/VSR1 data and S6/VSR2 data. The estimated computing needs were already reconsidered on the basis the 2009 CPU use until May 1st and of the new plans of the LSC-Virgo scientific groups which have since then evolved and become more precise (see VIR-030A-09).

The use until October 1st (about 90,000 kSPCint2k.day for CNAF/Bologna and 77.000 kSPCint2k.day in CCIN2P3/Lyon) shows a lower consumption with respect to the one estimated. We can thus lower our request for 2009 (see Table 1). For 2010, taking into account that some searches as been delayed – expecially the all-sky search for neutron stars ring-down on S5/VSR1 and S6/VSR2, which will be CPU energy demanding – we estimate for the need for 2010 will be the one reported in Table 2.
It has to be underlined that at CNAF/Bologna, in case of overloading, only 10% of the farm computing energy is dedicated to Group 2 experiments, so that mainly at the end of each year we keep experiencing problems in the Virgo jobs execution. This can limit also our capability of performing Data Analysis for the LSC/Virgo collaborations.

	CNAF/Bologna [kSPECint2k.day]	**IN2P3/Lyon [kSPECint2k.day]**
Continuous signals	200,000	0,000
Burst sources	5,000	100,000
Stochastic Background	5,000	0
Coalescing Binaries	70,000	50,000
Detector Characterization	5,000	5,000
Total	285,000	155,000

Table 1 : Computing needs for 2009

	CNAF/Bologna [kSPECint2k.day]	IN2P3/Lyon [kSPECint2k.day]
Continuous signals	200,000	0,000
Burst sources	10,000	200,000
Stochastic Background	5,000	0
Coalescing Binaries	100,000	100,000
Detector Characterization	5,000	5,000
Total	320,000	305,000

Table 2 : Computing needs for 2010

5.2 Storage

5.2.1 CNAF/Bologna and CCIN2P3/Lyon

The storage situation at CNAF/BOLOGNA is (October 1st):

- Disk: 258 TB total, 151 TB free - storage;
- Disk: 15 TB total, 6.2 TB free - user space.
- CASTOR: 20 TB total, 2 TB free – buffer disk
- CASTOR: 149 TB total, 16 TB free - tape

The storage situation at CCIN2P3/LYON is (October 1st):

- XrootD cache: 102 TB;
- HPSS: 250 TB,
- SRB cache: 25 TB

The increase of storage requested for 2009 is confirmed whereas, for 2010, the request has increased with respect what previously estimated because we plan to have 8 months of full data taking instead of 6 as previously stated and because of the increase in the data acquisition rate. See Table 3 for the requested increase in storage for 2009 and 2010. Note that the request takes into account also the possible free space remaining from 2009 at CNAF/Bologna.

Period	CNAF/Bologna disk and Castor [TByte]	IN2P3/Lyon XrootDd cache and HPSS [TByte]
2009	82 106	44 190
2010	185 20	245 276

Table 3: increase of requested space for storage in the Virgo CCs.

5.2.2 EGO/Cascina site

In Cascina, for the development of the monitoring and of the on-line analysis, it is foreseen the need of **4 TByte**, one of which with direct access to the web.