The Computing Model for Advanced Virgo (AdV) described here has been written taking advantage from the experience gained so far with the data taking and analysis from the first engineering runs to the latest Science runs VSR1-VSR4 (the last run ended in September 2011). It takes also into account the technological progresses of these years, from the original Virgo Plan, which is dated back to the year 2002, VIR-PLA-DIR-7000-122 [2]. The Computing Model reflects also needs and constraints arised from the LIGO/Virgo agreement[3], with which we have been facing during the last years, and have finally been able to address in an organized way in this Computing Model.

This document is intended to be a living document, following a regular schedule. A natural schedule could be given by the Engineering Runs, a sequence of test runs scheduled from January 2012 up to Dec. 2014, aimed to test different aspects of our activities towards "Advanced Detector Era" (ADE). Obviously, there are projects for which this will not be enough and will need to be finally tested and refined on real data. Being a "Computing Model" it does not contain detailed technical, manpower and costs information, which are discussed in the "Implementation Plan", which is again a living document, revised as we gain experience on the different items in it. The two documents will be jointly updated, since any result derived by new proposed technical solutions (described in the Implementation Plan) will lead to a modification of the "Computing Model". A "Management Plan" document closes the loop, detailing the procedure to be followed to verify the evolution of the computing model and its implementation. The "Management Plan" is a very important (roughly) 2 pages document.

To explore gravitational wave (GW) physics with the Advanced Virgo detector the Collaboration aims to define a Computing model that fully supports *accessing* and *analyzing* the data. In general analyses run on real LIGO-Virgo data, more rarely on simulated data. Therefore, the goal of the computing model is to define a production and analysis system able to guarantee an easy and robust access to data and resources. For this goal data distribution and data access are crucial points.

Advanced Virgo has a hierarchical model for data production and distribution: different kinds of data are produced by the detector and firstly stored at the EGO site in Cascina ("Tier-0"). Two copies of the data sets are stored to the national Computing Centres (CC), CNAF (Bologna) and CCIN2P3 (Lyon) ("Tier-1s"). A sub-set of LIGO data are also copied directly to our CCs (this is different from the procedure adopted for Virgo) and another sub-set is copied to Cascina for "low-latency" analysis, which have the need to produce fast results (see below). Some data, from CNAF and CCIN2P3, are also moved to "Tier-2s" (institutionals, managed by Virgo members), "Tier-3s"(institutionals, not managed by Virgo members), "Tier-4s" (users workstations).

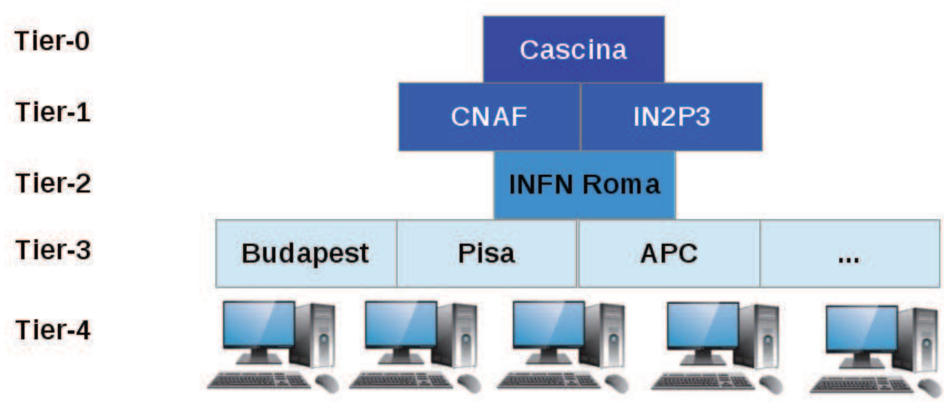

Figure 1: Virgo Computing Centers. Some analysis are carried on jointly with LIGO colleagues and thus also LIGO CCs are used, even if not shown here (we might consider them at the level of "Tier-3s").

The Cascina facility is dedicated during the runs to data production and to different detector characterization and commissioning analysis, which have the need to run "on-line" (with a very short latency, from seconds to minutes, to give rapid information on the quality of the data) or "in-time" (with a higher latency, even hours, but which again produce information on the quality of the data within a well defined time scale). The detector characterization analyses give support to both commissioning and science analysis. There is no permanent data storage in Cascina, and only data of a given time period (six months so far for Virgo, still to be defined for ADE) are stored there. The Scientific analyses are carried on at the Virgo Computing Centers (CNAF and CCIN2P3), with the only exception of "low-latency searches". And some analysis, due to the fact that we analyze data jointly with LIGO for many searches, are carried on in LIGO CCs. This as been detailed in the description of DA workflows. The AdV centres receive a copy of the data and provide storage resources for permanent data archiving. They must guarantee fast data access and computing resources for off-line analyses. Finally, they must provide the network links to the other Virgo computing resources.

The two CCs are integrated in the European GRID Initiative (EGI): for this reason we believe that pushing toward the adoption of the EGI products for ADE would be quite convenient. But the main constraint behond this is that we will always guarantee to users the possibility to work out of GRID, using local access to the data through "ffl" file lists, use of native batch systems and interactive when appropriate. Using EGI, we will take advantages of various tools and solutions already available and fully supported. To this goal, we have started working to implement a Data Transfer framework in the EGI environment.

However, some work is still needed to fulfill this scenario. For instance the data access interface used by Virgo at CCIN2P3 is based on XrootD; in order to be integrated in Grid it is necessary to install a specific layer between XrootD and the Grid Storage Resource Manager (SRM), and this has to be planned with the IN2P3 staff.

We also aim to run most off-line science analyses in the EGI framework. At the same time we need to guarantee that the developement, testing of new pipelines and the not computationally demanding analyses can be carried on outside the Grid environment, i.e. running interactive analysis or using local batch systems. This requires in particular the possibility to access subsets of data using "standard" technology, e.g. posix, scp, etc. To guarantee the accessibility of the data to a community of users as wide as possible we are thus investigating the implementation of a "user-friendly" Remote Data Access framework (based on a client-server communication protocol), which would be a layer between the users and the different underlying storage technologies. This framework would require the implementation of a "File Locator" database, which provides the information on the physical location of the data files and on the methods how to read them, and a database of the scientific metadata Clearly, this challenge (foreseen for the end of 2014) would take great advantage if the Tier-1 centers could adopt the same base technology (EGI).

Another hot topic is that as a consequence of the LIGO/Virgo agreement, the most of our Science analyses are carried on using both Virgo and LIGO data. Then in practice some analyses are performed in our Computing Centers and some others are carried on in ("LIGO Scientific Collaboration") LSC CCs. The latter are based on the "LSC Data-grid" (LDG) environment, whose workflows are not directly compatible with EGI workflows. It is therefore important, as we don't want to be in the position of not being able to run a particular analysis on our CCs (even only to repeat an analysis or to make additional tests), that the software we develop for our analyses should be independent of the used platform, or at least easily adaptable to different platforms.

This is a strong challenge, but we have shown it is possible. For instance, making use of the "Pegasus" workflow management system, we have already succeeded in executing the main CBC (Compact Binary Coalescence) search pipeline on both Virgo (at CNAF, in the EGI environment) and LIGO computing infrastructures.

We are also envisaging the possibility to run our pipelines in clusters using GPU technology. For this goal we started to translate part of our codes in the OpenCL language, which allows transparent execution on CPUs or GPUs.

A particular class of GW analysis are the so-called "low-latency" searches, which aim to provide fast alerts to the astronomical community in order to perform follow-up analyses of candidate GW signals. These searches require special solutions for both data transfer and computing workflow. The input data consist of the science data stream of the GW detector network (ADLIGO Hanford, ADLIGO Livingston and ADVirgo) and the data quality information. The produced output triggers are finally stored in a joint LIGO-Virgo database (GRACEDB is the one so far used) and sent to the astronomical community.

We foresee two different "low-latency" searches: the Multi Band Template Analysis (MBTA, a CBC pipeline) will run in Cascina; the Coherent Wave Burst (version 2) analysis (CWB-2G, a GW burst search pipeline) will run on one of the LIGO clusters.

The most important issues addressed by this model may be summarized as follows:

- guarantee adequate storage and computing resources at Cascina, for commissioning, detector characterization and low-latency searches;
- guarantee fast network links between Cascina and LIGO CCs for low-latency searches;
- guarantee reliable storage and computing resources for off-line analyses in the CCs (CNAF and CCIN2P3);
- push towards the use of geographically distributed resources (GRID), whenever appropriate;
- push towards a homogeneus model for Bulk Data Transfer (Virgo data from Cascina to CCs and LIGO data to CCs), Data Bookkeping and Data Access;