Computing and storage requests form for CSN2 experiments

Cherenkov Telescope Array (CTA) and MAGIC

1 Reference person(s) for computing issues

Name	INFN unit	Email	Cell. phone
Federico Di Pierro	Torino	federico.dipierro@to.infn.it	3453556193
Francesco Longo	Trieste	francesco.longo@ts.infn.it	

2 Description of the computing model

The INFN computing resources for CTA and MAGIC have been merged in 2019 to optimize their management and exploitation. In the following sections we will describe separately the MAGIC and CTA computing models, however there will be a single common request.

2.1 Data production, data transfer and data storage

Describe here your computing model, where data is produced, how and where it is transferred and where it is stored. Distinguish between CNAF (T1), Tier2 sites and other sites

MAGIC

MAGIC is a system of two imaging atmospheric Cherenkov telescopes located on the Canary Island of La Palma. They produce about 2TB of data per standard night, which are stored and processed firstly locally at the telescope cluster. As soon as the data are assembled and processed are sent in various format (raw, compiled, calibrated, etc.) to the PIC (Portal Informacion Cientifica, Barcelona, Spain) where they are stored in tape and disks.

PIC handles several activities for MAGIC: 1. Data transfer from the MAGIC site storage to PIC, 2. Transfer monitoring (for urgent analyses), 3. Raw data storage, 4. Raw data processing, 5. Data access, 6. Database, 7. Support, 8. Wikipage 9. Etc. Raw MAGIC data are roughly stored since 2009.

Data below corresponds to about 7 years of datataking.

TAPE	1.2 PB of raw data stored
GRID-DISK	152 TB
LOCAL DISK	95 TB

Beside real data, also Monte Carlo data are constantly produced for both standard analysis and peculiar scientific cases. Several computing centres are involved in the MC data production and analysis. INFN resources are not involved in the MAGIC data storage, nor in the official data reduction pipeline, nor in the Monte Carlo data production.

CTA

The Cherenkov Telescope Array (CTA) is the new generation ground-based observatory for very high energy gamma-ray astrophysics. To achieve a full sky-coverage the CTA observatory will be composed by two arrays of Imaging Atmospheric Cherenkov Telescopes (IACTs), one in each Earth hemisphere. The Northern array will be placed at the Observatorio del Roque de Los Muchachos on La Palma Island, Spain, while the Southern array will be located in Chile at the ESO site close to the Cerro Paranal. Each array will be composed by IACTs of different sizes to achieve a ten-fold improvement in sensitivity with respect to current IACT arrays over an unprecedented energy range extending from about 20 GeV to 300 TeV. The CTA project has entered its construction phase, the first Large Size Telescope (LST) is in the commissioning phase and the construction of 3 more LSTs is going to start soon at the Northern site. Several telescope prototypes have been built in different sites, INFN is highly involved in the prototype of the Dual Mirror Medium Size Telescope (SCT), inaugurated in January 2019 at the Fred Lawrence Whipple Observatory (Arizona, US).

The preliminary CTA Observatory Computing Model, at the end of the Construction Phase (foreseen in 2025), is extensively described in [1, 2], here we just summarize it: data will be produced by the telescopes in the 2 sites, compressed and distributed to 4 Data Centers. Computing and storage loads will be managed by the CTA Science Data Management Center. When fully operating the total data archive, including all raw data, MC, and several different data levels, will be 27 PB/y (6 on disks and 21 on tapes). The computing power needed for daily analysing data, data reprocessing and Monte Carlo productions, when CTA will be fully operative is foreseen to be ~14000 cores.

On a shorter term (2020-2021), before CTA Observatory operations will start, both commissioning and early-science data will be produced by the LST 1, the next LST 2, 3 and 4 and the SCT prototype (pSCT).

In 2020 gamma-ray data are expected to be delivered by the LST1 in La Palma. One night (10 h) of observation with LST1 will produce 90 TB of raw data, corresponding to an annual (~1300 h) raw data production of 12 PB. A data reduction of 97% of the acquired data is foreseen, this would reduce the annual storage to 400 TB. In 2021 the other 3 LSTs will be operational, increasing by a factor of 4 the requested resources. The current computing model for LST foresees the use of a powerful Computing Container (3 PB, 3000 cores) already installed at the site, where data are temporarily stored on disk, reduced and processed producing high level data. The raw data then will be transferred to other remote storage centers (PIC in Spain will be one, CNAF Tier1 could be the other one). The role of the remote data centers, beside the storage and preservation of raw data, will be also the data analysis of those data and the related Monte Carlo productions. The Computing container in La Palma, PIC and CNAF are already nodes of the CTA Grid Infrastructure, whose protocols will be used for data transfer.

Similarly, the computing resources will be handled via the middleware DIRAC [3].

In 2020 also the pSCT will start to provide data and CNAF could also be involved in those data storage and analysis.

Beside the real data coming from the telescopes, CTA will continue the massive production of Monte Carlo data, task where CNAF (as a Grid node of the VO.CTA) is already one of the main contributors [4]. In 2020 the Monte Carlo validation process will start.

2.2 Data Analysis at CNAF and in other computing centre

Specify here what data analysis is performed at CNAF and in other sites, comparing the computing energy used at CNAF with respect to the other sites

MAGIC data are analysed at PIC up to high level, users usually download those high level data (calibrated, etc) as well as Instrument Response Functions and Monte Carlo locally, to perform specific analysis. CNAF resources are used mainly to generate specific Monte Carlo simulations and for such high level individual user analysis.

The role of CNAF in the CTA final computing model is still to be defined. In case it will be proposed and chosen as one of the CTA Data Centers, it will contribute to scientifiic data storage and analysis. CNAF is already currently significantly contributing to the computing needs of the CTA Consotrium and to the INFN users involved in the project.

The main current activities are Monte Carlo productions and analysis, and CNAF is the second contributor in terms of disk and the fourth in terms of computing power to the CTA Grid Virtual Organization. Part of the allocated resources are used by INFN users, for specific productions and analyses.

In 2020 CNAF resources could be used to analyse LST1 data in cooperation with the La Palma IT container and PIC data center.

2.3 2019 outlook

Current status

The allocated resources for current year are going to be used as planned.

3 Pledge for current year (2019)

Including MAGIC resources that will be merged within the end of the year.

Experiment Name	Where		CPU [HS06]	DISK [TB]	TAPE [TB]
CTA	Tier1		5296	1265	270
	Elsewhere				

4 Proposed 2020 Pledges

Experiment Name	Where		CPU [HS06]		DISK [TB]		TAPE [TB]	
			Δ	New pledge	Δ	New pledge	Δ	New pledge
CTA	Tier1		0	5296	235	1500	130	400
	Elsewhere							

5 Motivations of the 2020 request

Free text, about 1 page for additional requests.

1. *Please comment the current usage of the computing and storage resources,*
2. *Detail if/why you are requesting additional resource and explain how the additional request is correlated to the current usage and to the evolution of the detector; quantify the request comparing to the expected detection rate, needed analysis, …*
3. *Insert the request in the framework described by the computing model*

The current (2019) resources granted by CNAF are well dimensioned for the CTA and MAGIC activities. We plan to fully use them within the end of the year.

In 2020, as detailed in the previous sections, the main computing activities involving INFN resources will be:

- MC productions as a Grid node of the CTA VO
- Storage and re-analysis of LST1 data. For this task we request, in addition to disk, tape and cpu resources, also manpower (0.1 FTE) to organize and to keep operational the file transfer, the web access and the software installation.
- Similarly it could be possible to store also pSCT data.

Considering the foreseen increase of the needs of storage we request 235 TB of Disk and 130 TB of tape, in addition to current resources.

6 Prediction of the computing needs in the following years

Try to estimate the evolution of your computing needs in the following years

In the next 2-3 years the number of telescopes in both Northern and Southern sites is going to increase, and with them also the data volume and the resources to analyse them will increase. The CTA Observatory has not yet defined a detailed computing model so it is not clear the context within CNAF resources will be included.
However accordingly to the existing preliminary CTA Computing model [1], considering some delay with respect to the plans when it was written, we can depict two extreme case scenarios for CNAF, i.e.: the case of CNAF being one of the 4 CTA Data Centers

and a minimal scenario where CNAF resources are exclusively dedicated to INFN users .

Scenario 1 (CNAF as one of the 4 CTA Data Centers)

The following predictions on CPU needs have been obtained considering the requirements of reprocessing one night of data in less than one day and to process the annual MC production in less than one month. The storage needs have been obtained considering the produced data volumes (different data levels, MC and Technical data).

In the period 2021 - 2028 the needed resources at CNAF would be:

Year	2021	2022	2023	2024	2025	2026	2027	2028
Cpu [HS06]	5500	5700	6200	7300	8700	10200	11700	13200
Storage [PB]*	3.2	7	10	12	15	18	21	24

*Raw data (DL0) will be stored on tapes, however after a staging phase on disk.
When fully operative, in the Data Center scenario, CTA will require to CNAF a constant disk space of 1.9 - 2.0 PB (all the rest on tape).

Scenario 2 (minimal)

In case CNAF would not be involved as an official CTA Data Center, the computing needs related to the activities of the CTA and MAGIC INFN communities (for instance: specific or alternate Monte Carlo productions, dedicated science case studies), would be of the order of 3000 HS06, 1 PB of disk and 150 TB of tape.

7 Comments

8 References

[1] CTA Collaboration, "Data Management Technical Design Report", Internal Document

[2] F. Di Pierro, "The CTA Computing model and INFN", presentation at INFN CCR Workshop 2017

[3] http://diracgrid.org/

[4] L. Arrabito, C. Bigongiari, F. Di Pierro, P. Vallania, “The Cherenkov Telescope Array”, CNAF annual report 2018