



Data Storage

# data storage

The total storage capacity available to CESGA users did not increase significantly in 2008 since the incorporations of Finis Terrae had been produced and accounted for in the prior year. The storage capacity on the 465.6TB tape was the only memory increase noted after the acquisition of 291 new LTO-4 tapes for the robotic library, reaching a total of 625.6TB. In total, available storage for users by the end of 2008 reached 1013 Terabytes which represented an increase of 57% with respect to the previous year.

The data storage service now uses a type of hierarchy for the data that is stored in order to assure the best quality of service as a function of the information (criticality of data and velocity of access). There were 77 requests for storage service of which 40 had an increase in their usage quota on the computing servers, 27 concerned massive data storage and 10 concerned security copies to disk (some users subscribe to more than one service).

## Storage service

### Criteria for the classification of information

In 2008 the Centre maintained the criteria introduced in 2004 regarding the classification of the types of data in the storage service.

This system permitted the accommodation of the different services to the specific necessities of each group of information, responding in this manner to the growing demand for quantity and quality of service such as the optimisation of the different storage options available in the Centre. A description of the 5 types of information that are available in the service is presented below in accordance with the classification previously defined.

**Scratch** has very low latency and maximum band width. It affects the computing production of the Centre. Regarding average capacity, the data are only stored for the duration of the execution of the computing jobs. Availability may be low (they are temporary data) and it is not necessary to make back-ups.

**Home Directories** contain critical data that are susceptible to being analyzed and modified at any moment. The functioning of the computing services of the Centre depends on their availability. As such, they should have (maximum) availability as a priority as well as an ideal balance between capacity (average, as a function of the number of users) and performance (average), of which back-ups are made on a daily basis.

**Massive data storage (MSS)** is utilized to store data bases and research results; normally the content does not vary (they usually are of the WORM type) and the access velocity is usually not critical, although they require a wide bandwidth to access the servers. Back-ups can be made according to demand.

**Back-ups to disk** are the copies of the data that users make with their own servers or PCs are stored in CESGA's systems in order to have a secure copy of their data. The availability of the service may be low. The service is offered by means of the network for which it is not necessary that the type of connection be one of high performance.

**Scratch Parallel** is very high performance (very low latency and maximum band width), similar to the first type with the addition that, in this case, the scratch data is shared between all of the nodes of the cluster and distributed among all of them). This permits an increase in both the access bandwidth to the files and the total capacity of the scratch well above the local disc capacity. Its availability may be low due to the fact that it depends on many components that are not redundant. No back-ups of these data are made.

### DATA STORAGE USED

TYPE OF STORAGE	CAPACITY (Gbytes)
Temporary or Scratch	133,480
Temporary Parallel	228,800
Permanent Disk	25,538
Tape	625,600
<b>TOTAL</b>	<b>1,013,422</b>

# DATA STORAGE RESOURCES 2008

