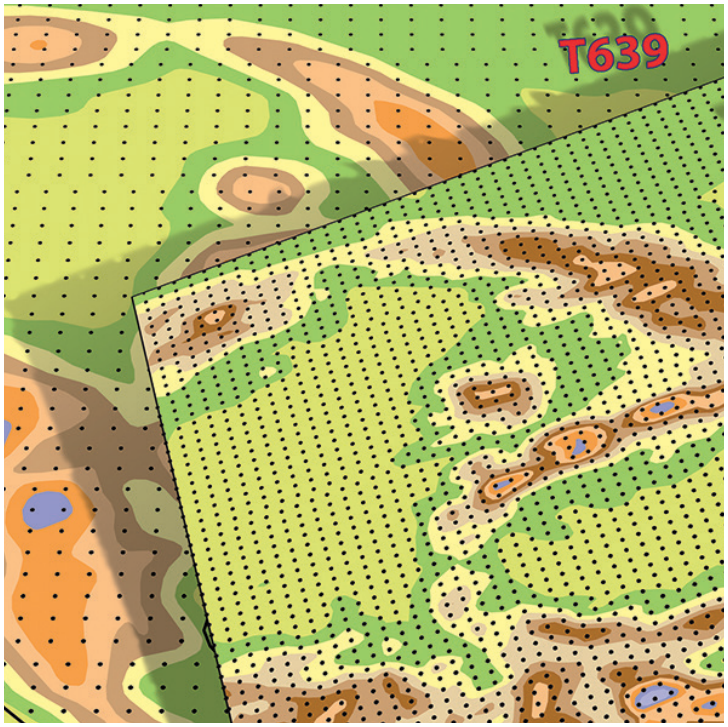
ECMWF Feature article

from Newsletter Number 124 – Summer 2010

COMPUTING

The Data Handling System



www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/7lb4my6g

This article appeared in the Computing section of ECMWF Newsletter No. 124 – Summer 2010, pp. 31–32.

The Data Handling System

Neil Storer

Weather forecasting makes use of, and generates, very large volumes of data: observations, analyses and the results of research experiments, all of which need to be stored for long periods of time. This data represents a valuable asset and an incomparable archive of meteorological information from the past 30 years. It is used worldwide by researchers in meteorological and environmental studies, and is also available for educational and commercial purposes.

ECMWF's Data Handling System (DHS)

For many years ECMWF has operated a dedicated Data Handling System (DHS) in which all ECMWF users can store and retrieve data needed to perform a wide variety of research and development activities. The ease with which data in the DHS can be accessed is regularly commented upon by visiting scientists and other users of the system.

IBM's High Performance Storage System (HPSS) is the underlying data management system in which all of the data in the DHS resides. However users do not use HPSS directly, they access the data via one of two applications, MARS and ECFS, that have been developed by ECMWF (Figure 1):

- MARS (Meteorological Archival and Retrieval System): a unique resource which allows users to access and retrieve a wealth of meteorological data via a meteorological interface.
- ECFS (ECMWF Common File System): a facility which allows users to store data that is not suitable for storing in MARS.

At the middle of 2010, 15 petabytes of primary data existed in the DHS (Figure 2), residing on approximately 18,000 IBM 3592 tape cartridges located in five STK powderhorn silos, along with one hundred IBM tape drives that processed almost 10,000 tape mounts per day. In addition, about 5 petabytes of secondary copies of the most important data were stored on LTO tapes in ECMWF's disaster recovery system. The DHS provides an excellent service to the users and equipment (such as servers, disk subsystems and tape drives) is added to the configuration each year to ensure that this high level of service is maintained.

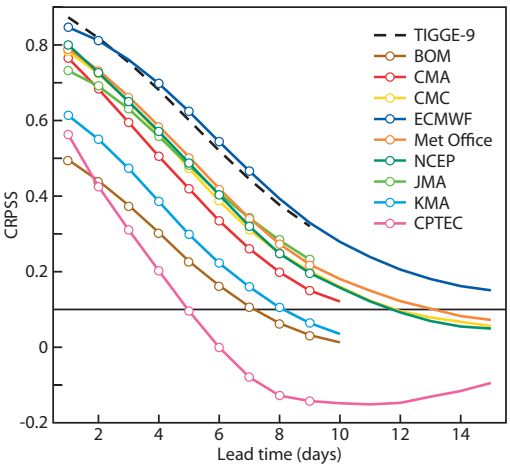


Figure 1 Examples of the parameters and commands used for MARS and ECFS retrievals.

doi:10.21957/7lb4my6g

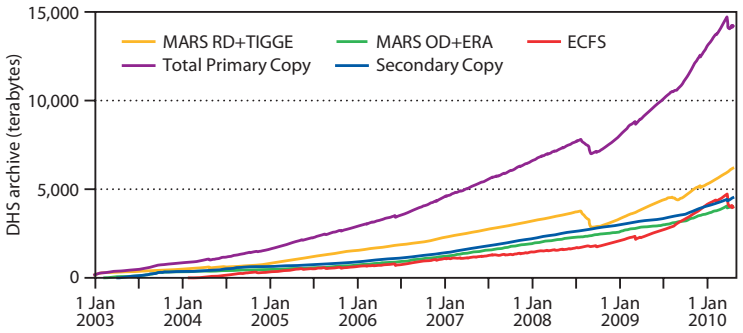


Figure 2 Historical growth in the data archive. MARS RD+TIGGE: Research Department and TIGGE data held on MARS. MARS OD+ERA: Operations Department and reanalysis data held on MARS. ECFS: Data held on ECFC. Total Primary Copy: Total primary data held on MARS and ECFS. Secondary Copy: Copy of Operations Department, reanalysis and e-suite data on MARS, and some ECFS data.

New DHS hardware

In the first half of 2009, the evaluation of a competitive tender was carried out for the replacement of two core components of the DHS: the IBM enterprise class tape drives and the StorageTek silos, which, having been in service for nearly 20 years will cease to be maintained at the end of 2010. A contract was concluded with Sun Microsystems Ltd. (now Oracle Corporation) in June 2009.

The replacement system will be installed in yearly phases and will take over completely from the old system later in 2010. Two SL8500 automated tape libraries (ATLs), 90 Sun T10000B tape drives and 6,000 1 terabyte tape cartridges were installed in October 2009. A third ATL, cartridges and over 60 more tape drives were installed in January 2010 and more equipment will be added in the third phase during 2011. Each SL8500 can house up to 10,000 cartridges (Figures 3 and 4).

Within each ATL eight small robots, known as ‘handbots’, move cartridges between the slots in which they reside and the tape drives (Figure 5). There are four sets of rails running along the inside wall of the ATL and each rail has two independent handbots that move along it. The two corridors within the ATL are connected at the far end in a horseshoe shape, which allows a handbot in one corridor to move all the way along that corridor, past the tape drives that are located at the far end, then into the other corridor.

Handbots can only move horizontally, along the rails, not between them. Should a cartridge, retrieved by one handbot, need to be mounted on a drive serviced by another set of handbots then the one doing the retrieving must bring the cartridge to the end of the rail furthest from the tape drives, where an elevator system is used to move it vertically to the rail servicing the relevant tape-drive. A handbot on that second rail then can mount the tape on the drive. It is possible for cartridges in one ATL to be mounted on drives that reside in a different ATL via the use of ‘pass-thru ports’. This helps to reduce queuing time when all of the drives in one ATL are in use but some are free in another one.

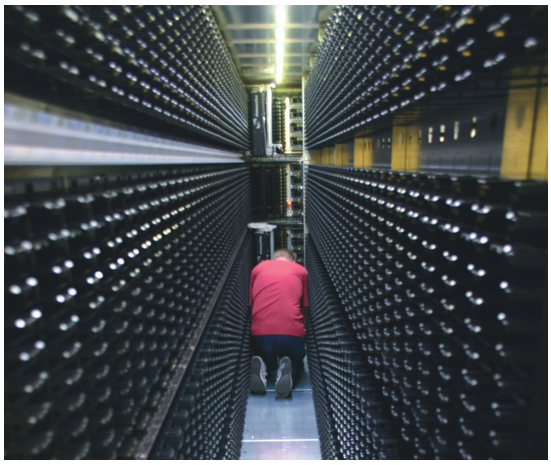


Figure 3 Inside one of the two corridors of an SL8500. Each ATL can house up to 10,000 tape cartridges.

Figure 4 Construction of the Sun SL8500 ATLs.

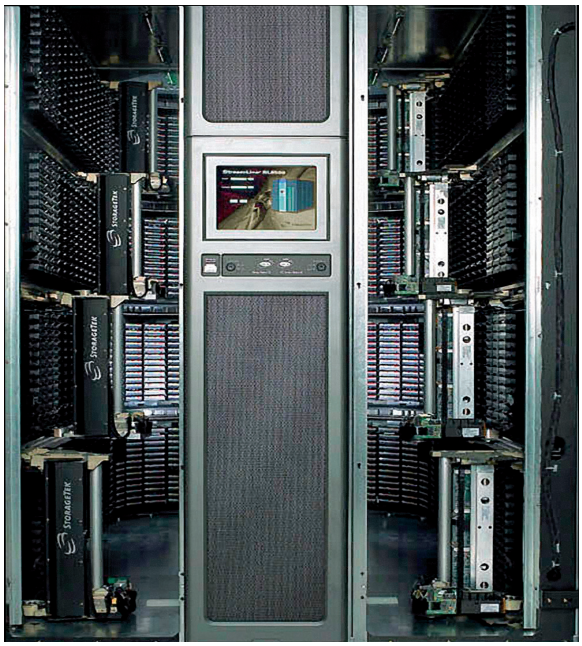


Figure 5 The eight 'handbots', four in each corridor and two on each rail.

Migration of data to the new system

Throughout 2010 all of the data on the IBM 3592 tape cartridges residing in the old STK silos will be migrated into this new system. This will be a major exercise since it has to be done carefully, ensuring that the data is copied across correctly, without impacting unduly on the quality of service provided to the normal users of the DHS. Once the migration process has completed, the IBM tape drives and the STK silos in which they reside, will be decommissioned and removed from the site.

In January 2010, ECFS and that part of the MARS service that deals with operational data (MARS-OD) started to use replacement system. At the same time the old MARS-OD data started to be migrated across. This migration completed in March. The migration of ECFS data was delayed until version 7.3 of HPSS was put into production as this has a new file aggregation feature that greatly assists the migration of this type of data. By the end of April 2010 all of the MARS data was migrated.

The STK silos have stood ECMWF in good stead for twenty years. It is hoped that the new ATL complex will prove to be as reliable for the next twenty.

doi:10.21957/7lb4my6g