DATA MANAGEMENT IN THE AUSTRALIAN NATIONAL METEOROLOGICAL OPERATIONS CENTRE

C. W. Sanders
Bureau of Meteorology
Melbourne, Australia

1. OVERVIEW

The Australian National Meteorological Operations Centre (NMOC) has similar data management problems to most other centres who run global models. These are to store, in a readily accessible form, large amounts of observations, model output fields, and various other data.

In NMOC we store almost all observations, some overseas model output fields, and most local model output fields in a NEONS/Oracle database. In general, ten days of data are kept, although this can be varied if needed. We moved from an Empress database to Oracle a couple of years ago, in order to standardise on one RDBMS. NEONS/Oracle is being installed in the regional offices as part of the AIFS (Australian Integrated Forecasting System) installation, while the National Climate Centre also uses Oracle.

A significant amount of data is copied to McIDAS (Man-computer Interactive Data Analysis System) MD (Meteorological Data) and GRID files. Our regional offices make extensive use of McIDAS and there are many McIDAS users in our Head Office as well. Most are using version 7.3 of McIDAS, although there is some usage of other versions. There are plans to alter McIDAS to read directly from NEONS, and if this happens, the copies will become redundant.

The output from our locally run models GASP, (Global ASsimilation and Prognosis), LAPS (Limited Area Prediction System), TLAPS (Tropical Limited Area Prediction System), and MESOLAPS is generated as NetCDF data files. In general only the last one or two runs are held on-line in this form, with the database and SAM-FS (the current archive system) being used for older data.

Satellite data is held in McIDAS AREA files. There are currently no plans to change this. Some data have not yet been integrated into the database. In particular, some observational data are held in McIDAS MD files but not in the database.

SAM-FS is used for archival of data. Most observational data and overseas model output are archived in NEONS dump format. Most output from locally run models is archived as NetCDF files. A variety of other formats are also used.

2. DISTRIBUTION TO USERS

Most of the output from the locally run models is currently distributed to the regional offices in the form of McIDAS GRID files. This method is also used for some of the overseas model output, which is received over the GTS in GRIB code, while a limited set of observational data is also distributed as McIDAS MD files. Collection of products from the central data server using McIDAS ADDE (Abstract Data Definition Environment) is also available. This is becoming the preferred approach, as it minimises communication overheads if properly used. Currently, some of the less used data, and some large NWP outputs where only a small subset is required routinely, are distributed using McIDAS ADDE. As the regional offices install AIFS, which also uses NEONS/Oracle, Oracle networking will be considered as the means of giving the regional offices direct access to some of the data held in the head office database.

In addition, the output from GASP is distributed over the GTS in GRIB code. LAPS and/or TLAPS may be added, although there are no firm plans to do so yet. Much model output, both from the locally run and from the overseas models, is also distributed as pictorial charts to a wide variety of customers, both inside and outside the Australian Bureau of Meteorology. These charts are usually distributed by using various forms of facsimile, although a few are distributed as image files.

Selected products (usually either pictorial charts or text files) are placed on a World Wide Web server (www.BoM.GOV.AU). Some of these are publicly available, others are restricted to registered users. Limited

use is made of an ftp server (ftp.BoM.GOV.AU) to make some data available generally (via anonymous ftp) or to registered users. This is in various formats, depending on the needs of the users.

3. NEONS/ORACLE DATABASE

NMOC uses a fairly old version of NEONS, (Version 3.5 with some 3.6 enhancements and a few local enhancements). We hope to upgrade this in the longer term. The NEONS database is used for observations and for model output fields. It is not used for satellite pictures.

Data received over the GTS or over the BoM's internal communications system (CMSS, Computerised Message Switching System) are handled by using a transaction driven system. The incoming messages are normally passed to the appropriate decoder, decoded, and stored immediately in the database.

The output from the locally run models is loaded into the database by using batch updates, which are controlled by our scheduling system (SMS, Supervisor Monitor Schedular, obtained from the ECMWF). The batch updates are initiated when the model reaches certain pre-determined time-steps. Currently the GASP output is loaded after every 24 hours of model time, and the LAPS/TLAPS output after every 12 hours of model time.

Locally processed TOVS data are loaded into the database by using tasks scheduled by SMS. At regular intervals (currently every 15 minutes), a check is made to see if any new data has arrived. If any has, it is loaded into the database. A similar system will be used for locally derived cloud drift winds.

Some minor observation types, including PILOT SHIP, SYNOP MOBIL and a number of others are not yet decoded or stored in the database.

4. DATABASE PERFORMANCE

We are generally happy with the performance of the NEONS/Oracle combination. It is far superior to the NEONS/Empress for observational data, although it is inferior to NEONS/Empress for the model output fields.

No detailed investigation of the reason for this has been done, but Oracle seems to be optimised for short record lengths. The observational data have typical record lengths of 50 to a few hundred bytes, while the record lengths for the model output are much longer, typically about 10 to 20 kbyte. For short record lengths, Empress seems to do one disk write per index or table update, while Oracle appears to write one block (8 kbyte) at a time, giving a large performance gain. For long records, Empress still seems to do one write per index or table update, while Oracle appears to split the record across multiple blocks, does multiple writes, and does not perform as well.

The performance of the NEONS/Empress combination on observational data was barely acceptable. Delays of up to an hour or more were common at peak observation times (shortly after 00 and 12 UTC). An outage of one or two hours would result in a backlog that would take more than 6 hours to clear. On the other hand, the performance of NEONS/Empress for model output fields was good.

The NEONS/Oracle combination is much faster at handling observational data. Significant delays are very rare, with the delay between an observation arriving and being stored in the database usually much less than a minute. The backlog created by an outage of one or two hours normally clears within a few minutes. On the other hand, the performance for model output is worse than it was with Empress, although it is acceptable. A fair amount of work on scheduling, re-assigning tasks and other similar enhancements had to be done to achieve acceptable performance when batch loading large numbers of model output fields.

Another problem that surfaced during the Empress to Oracle conversion was that Oracle requires much more disk space than Empress for the same data. We were aware that Oracle had a reputation for being "Disk Hungry", and had allowed for an increase by a factor of one and a half to two. The actual increase was closer to a factor of three. Oracle appears to be particularly inefficient at storing the model output fields.

Again, no detailed investigation has been done, but Oracle seems to be unable to efficiently allocate large records to blocks. Oracle seems to waste space if a large record, plus the storage overhead (a few bytes) is not a multiple of the block size. It seems to be particularly bad if the record size is slightly more than a multiple of the block size.

5. DATABASE BACKUP

Backing up of the database is a major problem. Normally, a database is taken off-line and shut down while the backup takes place. This is not feasible for an operational database, as we cannot afford regular (daily or twice daily) outages of six hours or more. In addition, the last 6 to 12 hours of data are the most critical, so conventional backups are of limited use. One advantage we have is that the communications system maintains the last two days of traffic, and this can be used (via a replay facility) to recover some data after a failure.

It was judged that rare (after a catastrophic failure) outages of 6 to 12 hours could be coped with. After a lot of discussion, the chosen strategy was to make regular backup copies of the static and slowly changing data using Oracle tools. This includes the database structure, station dictionaries, parameter descriptions and other similar data. The volume is of the order of a few Byte. The actual data (observations and model output) is backed up by making regular dumps from the database in NEONS dump format. These files can be easily loaded into the same or another, compatible, NEONS database.

Recovery after a catastrophic failure then involves building a new database, loading the static/slowly changing data, and the most important (most recent) of the NEONS dump files, starting the decoders and re-playing traffic. Tests have shown that this takes about 6 hours to get a database capable of supporting the main functions of NMOC. The remaining data can then be loaded at leisure from the NEONS dump files.

6. ARCHIVES AND OFF-LINE STORAGE

The Australian Bureau of Meteorology has recently replaced an archive and off-line storage system based on Epoch software with one based on SAM-FS. SAM-FS runs on a Sun E4000 system with 120 Gbytes of raid disk, two StorageTek 9710 library storage modules and twelve DLT7000 tape drives.

A major problem is that there is no "universal" interface to access all of the archived data, the on-line data in the database and the on-line NetCDF files. We hope to install MARS and connect it to on-line NetCDF files, NEONS and the SAM-FS archive to provide this uniform interface.

SAM-FS has proven to be highly flexible and configurable. Different categories of data can be treated differently, based on directory and file names. Each data category can have different retention periods and numbers (up to a maximum of four) of tape copies maintained. Provision can be made for off-site disaster backups, for example, three tape copies of the analyses are maintained, one of which is an off site disaster backup. There can be other differences between categories of data, and data from a given category can be restricted to a particular set of tapes so that related data is held on the same tape or set of tapes.

In addition, SAM-FS also automatically manages disk space, copying files to tape and restoring them as needed. It also allows data to be retained on disk for a specified period, which is useful if it is known that the data will be heavily accessed during the first few days after it is created.

We have configured three broad categories of data

- Permanent archive, kept forever. These are mostly the analyses and observations. There is also provision for individual users to permanently archive limited amounts of data, such as special case studies.
- Medium term, kept for four years. These are mostly the model forecasts.
- Short term, kept for 3 months. One example of this category is the communications messages and communications logs.

SAM-FS also stores the data on tape in unix "tar" (actually GNU tar) format, which means that the data can still be recovered, even if the SAM-FS directory structure and indexes are destroyed. Our previous archive

system, Epoch, used its own, private tape format, and if the database of which files were where was lost or damaged, the data was effectively impossible to recover.

The performance of SAM-FS is far superior to Epoch. Whereas recovery of a large file using Epoch took 6 hours on average, and occasionally took 12 to 24 hours, using SAM-FS it is typically 1 to 3 minutes. Despite its very poor performance when recovering files, Epoch's performance at saving files to tape was not too bad, giving rise to many comments about a write-only archive.

7. FUTURE CHANGES

Improvements we hope to make in the reasonably near future include

- Add the remaining observations (PILOT SHIP etc) to the database
- Change the distribution system to the regions. Currently most data is sent to the regions whether they need it or not. We are in the process of changing less used data to a "send on request" system to reduce network traffic. Heavily used data will still be sent automatically.
- Change as much archived data as possible to NetCDF or NEONS dump format, eliminating most of the various ad-hoc formats currently used.
- Install MARS to provide a uniform data access method.
- Improve access to our data holdings by external users, mainly by using the WWW, but also making use of ftp, anonymous ftp and other suitable methods.

We have no firm plans for other major changes. One thing we would like to do, if resources permit, is to upgrade to a more recent version of NEONS. The version used by the Meteo France seems to be significantly better than ours.

Another possible future improvement is allowing direct access from the World Wide Web to the database, using CGI scripts, Java, Oracle's WWW interface or some other suitable method.