A NEONS-BASED WORKSTATION FOR THE REGIONAL FORECAST OFFICE

Don Dueck,
Jean Brunet
Atmospheric Environment Service
Downsview, Ontario, Canada

Neil McLennan
Atmospheric Environment Service
Vancouver, British Columbia, Canada

1. INTRODUCTION

The Navy Environmental Operational Nowcasting System (NEONS) database as described by Jurkevics et al (1990) and Shaw et al (1990) provides for a simple, compact yet flexible database design that can accommodate a wide variety of environmental datasets. It was initially developed to support environmental research and the Naval Research Laboratory (NRL). The requirement for the Atmospheric Environment Service (AES) of Canada was quite different. AES required a database that was capable of supporting the real-time operations of a regional weather centre. NEONS had to potential to fulfill the database needs of such an office but modifications and testing would be required to confirm this possibility.

This paper describes the requirements of a database to operate effectively in the environment of an AES weather forecast office, the changes that were required to NEONS to meet those criteria and how NEONS is currently used within AES.

1. REQUIREMENTS

The amount of data that flows into a regional weather centre has undergone a significant increase in the past 5-10 years and shows no sign of abating. The demand for satellite imagery, radar data, numerical model output, meso-scale surface data and other less traditional data types from forecasters is expanding as they seek to provide better services to their clients. This explosion in the volume and variety of data has swamped the capability of the current environmental databases used within AES. When the design of advanced meteorological workstations for AES was just beginning in 1990, it was clear that a new database was required to manage the volume and variety of data sets.

The requirements of any database that was to fit into this real-time environment were identified as:

1) Performance
2) Geometry management
3) Reliability and maintainability
4) Flexibility
5) Adherence to accepted standards

Performance, which is the most important criteria, had two facets. The database had to support the continuous, simultaneous ingest of data from a multiple of sources of data. The data streams were to include satellite data from the Geo-stationary Operational Environmental Satellite (GOES) and the NOAA polar orbiting satellites, numerical weather prediction data from the Canadian Meteorological Centre (CMC), radar data, sferics data and surface and upper air meteorological data. Concurrently the database was to support a number of simultaneous users, each of whom was to retain an 'interactive' feel in their application.

The volume of data arriving in a typical weather offices is such that to view each data set independently and then make the necessary correlations between them is virtually impossible. The capability must exist to view them in an integrated manner to allow the meteorologist to make an accurate diagnosis of the situation and to reject any erroneous data values. To meet this requirement, the database has to manage the geometry of each piece of data so that the different data sets can be super-imposed.

The database must be flexible and extensible. The variety of data sets is rising and every one must be incorporated into the database so that applications can access them with only minor modifications. This requires a flexible data schema which will allow the addition of new data types and/or parameters and a consistent programming interface that permits the input and output of data through generic functions.

There are a number of special reliability considerations when running a database in an operational environment. The data flowing into the weather offices occurs continuously 24 hours per day and 365 days per year from a multitude of different sources. Ingest must be reliable as if it fails then a piece of data, which might be the crucial piece of the meteorological puzzle, is lost as there are no repeat transmissions. Data is only useful for a limited time so the database ingest algorithms must monitor the valid time of each piece of data and delete or archive any expired data automatically. This function will also maintain the database at a more or less constant size and reduce the amount of maintenance and investment in disk drives. In an operational environment, there is no scheduled down-time. Any maintenance or clean up activities must occur while the database is active. This puts significant constraints on the database schema and how it is implemented.

Meteorological data is exchanged around the world between member countries of the World Meteorological Organization. To expedite the inter-change of information, the database must support the transmission and storage standards of the WMO for gridded data (GRIB) and point data (BUFR). The database must also utilize current computer industry standards. This is likely to ensure a longer usefuly life span for the database than if it relies on special features of a particular hardware or software architecture.

This is not an exhaustive list of the requirements for a database to operate effectively in a real-time environment but only the minimum standard that it must reach.

2. DATABASE SCHEMA

2.1 Operational Requirements

The NEONS data schema as described by Jurkevics et al (1990) and Shaw et al (1990) has the capability to hold satellite imagery, two-dimensional gridded datasets and point data. In addition to the raw data, the sufficient information about the data was stored to permit fast and flexible retrieval of the data for specific geographical and date/time windows. However since NEONS was originally developed to support the research activities of NRL, it was not clear whether the original design would meet the different needs and priorities of a real-time environment.

The original design of NEONS was tailored towards a batch mode of ingest where a large volume of data was downloaded at a limited number of predictable times during the day. This not the model required for a weather office where data arrives in small volumes from a many sources 24 hours per day. A typical AES forecast office might receive 200 satellite images, 3500 grid fields, up to 50,000 lightning reports, 2500 mesonet reports, 40,000 land station reports, 4000 ship reports, 1500 upper air reports, and 700 radar products in a single day. Some data arrives from local sources, some via modem from other municipal and provincial government organizations, some a wide area network from the AES super-computer centre and others via asynchronous downlink from satellite.

The other major requirement of a weather centre is that the data be available very soon after it arrives. It is not acceptable to have more than few seconds or so delay from when the data arrives and when it can be accessed by applications. The dataset concept of NEONS did not support this requirement as data cannot be accessed in the primary tables until the dataset is closed and the associative realm updated.

2.2 Schema Changes

To meet these requirements the method by which data was ingested into and retrieved from NEONS was altered. Each data type has associated with it any number of primary tables. The data manager can decide the volume of each data type to store depending in the number of these tables that are created thereby controlling the over all size of the database. These tables are used to store the incoming data in a cyclic manner with each table holding a fixed number of hours worth of data. When all the tables in the set are full, the 'oldest' table in the set is archived, emptied and then re-used for ingest. By this method the database size is constant, retrieval is fast as each primary table is small, the slow 'delete' command is avoided and archiving becomes automatic.

To accommodate this ingest scheme a number of changes were required in the schema. Attributes were added to indicate the number of hours of data each primary table would hold. Also, five attributes were added to track the times of the data within the various primary tables. The possible values of the existing status attribute were also expanded to include:

1) Empty - the primary table is empty and will not be used
2) Load - the primary table can be used for ingest
3) Full - the primary table is full and ingest is not allowed
4) Archived - the primary table to full but ready to be emptied and re-used
5) Current - the primary table is not full and is the table currently receiving data
6) Read-only - ingest of archived data only is allowed

Using the status and the time information the ingest routines are able to determine which table and which dataset to write new information into. The arrival of a new piece of data can trigger the closing of a dataset and/or the closing of a primary table. In the former case, a pointer is updated and a new dataset started. In the latter case (which is always accompanied by a dataset closure), the 'oldest primary table is emptied and its status is changed to 'current'. Ingest will then switch to the re-cycled table.

One additional change was made in 1993 when the 'llt_val' attribute was found to be inadequate as a station designator. It was impossible to design a packing scheme to allow more than 5 alphanumeric characters into the long integer attribute. This presented severe difficulties for the ingest of climate data and 6 or 7 letter ship identifiers. A new character attribute was added to the LLT primary tables to overcome this shortcoming.

Other minor additions were made to the data schema to support particular applications with the weather office. For example, two additional attributes were added for imagery. These contain X and Y offsets required to correctly geo-locate GOES imagery. The offsets are derived interactively by the user as the image is manually panned to register it with a basemap and then stored into the database for use by other applications.

2.3 Programming Changes

The NEONS routines for ingest and retrieval were both altered to take advantage of the data schema changes.

The ingest routines were altered to utilize the new attributes in the schema.

The read routines provided with NEONS scan the associative realm for a list of datasets and primary tables that contain data that match the date/time query. However the NEONS ingest routines do not update these entries until the dataset is closed, so any search will bypass the open datasets. This prevents the users from accessing data in those datasets. Since real-time users must see the data immediately the read routines were changed to bypass these entries. An index using the epochal time was placed on the primary tables and retrieval was done by examining the list of primary tables that contain data with the requested date/time window. The data is then read directly from the list. The retrieval rate was much higher using this method; however, for massive primary tables this statement may not be true and the original NEONS output routines might be superior.

NEONS handles imagery in a very flexible manner. The ingest and retrieval algorithms permit an arbitrary number of bits per pixel but the penalty is a loss of performance due to the extra calculations required to pack the imagery. Satellite imagery used within AES operations is restricted currently to eight bits per pixel. New read and write functions were written to optimize performance based on this knowledge.

3. OPERATIONAL IMPLEMENTATION

3.1 Current Status

Benchmark tests described in Tran (1991) confirmed that the adapted NEONS met all the requirements listed in section one. Basic performance tests were conducted using a number of scenarios that would simulate an operational weather office. Various modes of ingest in combination with computationally intensive (e.g. objective analysis) and interactive applications were tested. The ingest functions were unaffected by the extra load imposed by the applications indicating that the limiting speed is the rate of arrival of data not the database. Users did not notice any significant loss of interactive 'feel' during periods of heavy database activity. Using these results, NEONS was accepted in 1992 as the foundation for the new graphics tools to be built.

Operational implementation of fully functionally NEONS databases and associated graphics software began in 1993. Three regional weather offices are now using NEONS extensively with plans to install the system at the remaining six offices in 1994. The databases are ingesting about 250 megabytes of information a day of image, grid, and point data (all the types described in section 2.1 except radar imagery). Three to four days of data are kept on-line at any given time resulting in a constant database size of 1 gigabyte. Expired data is either automatically rolled to a secondary device for long term storage if archive is desired or discarded if the archive system is disabled.

3.2 Future Plans

Now that NEONS is moving from a pure development environment to full implementation the flexibility and expandability of the schema will be tested. Each office has unique data sets that must be added to the database.

A number of data sets are absent from the NEONS schema including climate data, radar, and video.

Climate data can be handled as an LLT type but the challenge here is the volume of data. For British Columbia alone, climate information would occupy 8-9 gigabytes of disk space. The ingest routines for this data have be altered to compress the packed bitstream before writing to the primary tables. This has resulted in about at 30 percent saving in disk usage. This technique will probably be applied to other LLT types to maintain the generality of the I/O routines.

Radar can be considered as a three dimensional field and the new NEONS 3D grid type could be used but this may not necessarily be acceptable as it would compromise some of the descriptive and associative tables. Consequently some design work has been done on the most desirable methods of saving radar information but no substantive work has been done on the schema.

AES operates a number of video cameras at automated weather stations which supplement the normal reports. The video images are collected via modem and placed into the file system. Another extension to NEONS is planned to accommodate these JPEG images.

Eventually, the long-term goal is to have all types of meteorological data used by the forecaster available through NEONS-type calls, whether it is stored locally or distributed over our wide-area network.

4. SUMMARY

The NEONS has proved to be a very capable database system for supporting multiple interactive users in an operational real-time environment even when ingesting a wide variety of data concurrently. To meet It requires little maintenance, is compact, flexible and fast. It was an excellent choice to fulfill our requirements as an operational meteorological data server.

5. BIBLIOGRAPHY

Jurkevics, A., R. Titus, and J. Clark, 1990, Environmental Database for the Naval Environmental Operational Nowcasting System. Pre-prints, AMS 6th Conf. on Interactive Information Systems for Meteorology, Oceanography and Hydrology, Anaheim Ca., 80-83

Shaw, C., E. Schwartz, and Ted Tsui, 1990, Design of the Naval Environmental Operational Nowcasting System, Pre-prints, AMS 6th Conf. on Interactive Information Systems for Meteorology, Oceanography and Hydrology, Anaheim Ca., 233-236

Tran, S., 1991: WSO Prototype Database System Test Document. Unpublished Manuscript