BDCLIM : THE FRENCH CLIMATOLOGICAL DATABASE

F. Bénichou
Météo France
Toulouse, France

1. INTRODUCTION

One of the charges of the Climatology and databank national Service, located in Toulouse is the management of the Bdclim, the National Oracle Climatological database.The Bdclim was born in 1989 to manage the last two climatological data years on a dedicated computer , a CDC930 computer under NOS/VE. The service was so pleased with the use of RDBMS that it decided to replace the computer and enlarge the database. Since january 1993, the bdclim has been running on a CD4680 (128 Mbyes RAM, 100 simultaneous users, one processor,17 Gbytes magnetic disk), under unix and Oracle V6 (the kernel with the SQL interface and the SQL*FORMS application generator).

The Bdclim intends to manage on the one hand , conventional data from the 7 upper air sounding stations, the 168 synoptic stations, 700 automatic stations and 4000 climatological stations (beheld by volontary people and measuring just extreme temperature and daily rainfall), and on the other hand metadata such as station idents, the story of changes in position, data producers, type of stations , frequency of measurements and catalog of data.

The most frequently used data are on-line , that is to say :

- raw data
  - data from the synoptic stations since 1949 (daily and 3-hrly parameters)
  - data from the long climatological series (230000 years-stations)
  - all data from the current year (synoptic, automatic and climatological stations)

- computed data
  - monthly data (200 synoptic stations) since 1949
  - decadal data (200 synoptic stations) since 1949
  - normals 1951-1980 and 1961-1990
  - daily and monthly records (200 synoptic stations) since 1949

These data represent about 150 oracle tables and 12 Gbytes. The biggest table contains 6 millions rows for a size of 650 Mbytes.

2. THE RELATIONAL TABLES

Five types of raw data tables have been defined :

- a daily table for the daily observations of the synoptic or automatic stations including daily radiation observations
- an hourly table for the hourly data of the synoptic or automatic stations .
- a sounding table for the upper-air sounding observations
- a climatological table for the daily rainfall or temperature data of the climatological station
- a precip table for the intervals of precipitation with constant intensity

Monthly , decadal, record and normals tables are the main types of computed data tables defined for the synoptic stations :

For the sake of performance and physical spaces, we have been led to define several tables for every type of tables. Thus, the hourly long series data are split into 7 tables, one table per regional center.

The structure of the records is homogeneous for all types of tables . It is composed of :

station number, date of observation, list of parameters

Two oracle indexes have been created, one on the key of the relation (station number + date), the other one on the date.

3. THE ACQUISITION OF DATA FOR THE BDCLIM

The acquisition of the hourly, daily and precip data is performed every day after the collection of ascii files sent by the 7 regional centers through a private X25 network. Those of the climatological data is similar but works just once a month. The use of floppy disks appears just for the acquisition of the upper air sounding data, once a month. In all cases, the data loading is assumed by fortran or C programs . The sql*forms application generator or SQL commands are used to update or visualise data.

Controls of data are made in the regional centers. These centers have the same database as the BDCLIM, but are limited to the stations of the region and with a smaller length of time (just a few

months of data). They receive every day a special message, 'the Mescrq' with daily and hourly data from the departmental centers about the synoptic stations and the automatic stations. After loading data from the Mescrq into the oracle database , temporal and spatial controls are performed, the anomalies are sent to the stations which have to validate them. The original data and the quality codes are kept in tables. At the end of the afternoon, when the validations are received, some missing data are estimated, ascii files with all the data of the day or previous corrected data are sent to the national center. For hourly and daily data, it represents 6 files per region (for each type, one file for the final data, one for the quality codes and another for the original data).

4. SECURITY ASPECTS WITH THE BDCLIM

The security of the access to the database is well assumed by Oracle : user and password - 3 levels of users : the lowest level just authorizes the select of data, the middle one allows to create objects, the upper one has the database administrator grants . Limits of physical space may be defined per user. The owner of one table chooses the grants on the table he wants to give the other users (delete rows, update rows or columns, insert lines, just select rows). With the new version of Oracle (V7), it will be possible to create group of users with profile and with limits of number of sessions, CPU or I/O per session.

For the recovering after a crash , Oracle stores in real time, automatically all the changes which are made on the tables , in logfiles. For the bdclim, these logfiles are put each evening on a mass storage computer (a convex on unitree connected to 2 storage tek silos). Furthermore, a save on exabyte of the unix files which compose the entire database is done once a week . In case of a crash, oracle restores a clear database with the save files and the logs.

The BDCLIM has never problems of tailorage of date because Oracle puts locks on rows updated (two users cannot simultaneously update the same row and the users see data being updated in their 'before update status').
The most serious drawback of the present Version 6 of Oracle is that the constraints which are specified when the tables are created are not effective. So, many controls of data are done after loading. In a few months , this problem will be solved : on the one hand, these constraints will be effective and on the other hand, it will be possible to implement triggers (procedures automatically activated when an event appears .)

5. THE ADMINISTRATION OF THE BDCLIM

Two persons are fully busy with the BDCLIM administration especially to keep performances at a high level, which means : a good initialisation of oracle parameters such as size of data blocks, memory, logs , rollback segments, ..., a good definition of indexes and a survey of their use, an appropriate structure of the database (number of files, size and location of these files). Some audits are sometimes necessary.

6. THE DATA ARCHIVE

The less used data do not remain on line in the BDCLIM, especially data from automatic , upper air or overseas stations. These data are stored under ascii files compressed by the unix compress on the convex mass storage computer . To make easier the import of data in the Bdclim, it was decided to have one type of file per each type of table with a format like the matching table with the fields separated by a semi-colon : " numberstation;date;parameter1;parameter2; ...parametern" . Spaces and unitree constraints led to a definition of one file per station for the daily or hourly data and one file per station and per year for upper air data.

To make this archive transparent to the user, a software has been developed with unix commands and C programs, which demands to the user just to precise the type of table , the lists of stations and parameters he wants and the information 'ascii file' or 'table'. In this latter case, an oracle table will be created in the Bdclim, in the other case , an ascii file will be provided: The principe of this software can be summed up as below :

- elaboration of a file of directions
- sending of this file to the convex
- activation of the retrieval of the tapes containing the necessary files
- extraction of the good stations and parameters .
- gathering the results on one file.
- sending the result file to the bdclim computer with redo logs
- in the case of a table request, loading data in the bdclim (using the sqlload oracle tool)

7. CONCLUSION

The use of RDBMS in the climatological service is a success : access to data has became very easy and consequently the quality of data has been bettered, developement times have been much reduced, security has been greatly improved . Up to now, many sophisticated applications like graphical applications were done on PC or on other computers after retrieving data from the bdclim through ascii files. This operating technique is being changed : several workstations have been bought for the national climatological service, the SQL*NET tool (network tool) is being installed on the climatological server and these workstations and others workstations (research center ..), so in 1994 , the client-server architecture will be operational .