THE DIAPASON SYSTEM

J.Clochard
METEO FRANCE
Toulouse, France

1. GENERAL ISSUES

DIAPASON is the acronym given to the new automated production system of METEO FRANCE.

The previous data processing system was based since 1982 on CDC mainframes, running a proprietary operating system (NOS). One machine was operational and another one could act as backup, but there was a single set of disks. Limitations due to age were felt everywhere; in hardware of course, though machines had been replaced once: power, memory size... but also in software: no data bases, configuration split into a lot of (text) files, no or poor support of new standards (TCP/IP, PostScript, BUFR...) and/or terminal displays... Further more, maintenance had become more and more difficult, with knowledge concentrated on very few people.

So the goals of what was called the DIAPASON project were to:

- renew the data processing system of METEO FRANCE;
- move to a non-proprietary operating system: UNIX;
- give power for data management and product elaboration;
- improve reliability;
- reduce maintenance costs;
- lead to high inter-operability with other operational systems, up-to-date in terms of software and communication protocols;
- gives capacity to evolve and adapt: no strict vendor dependence, use of standards.

Nearly nothing could be reused from the previous system, except tasks running on CRAY (Numerical Weather Prediction and some related jobs). So we started a complete analysis, based on Object-Oriented approach (OOA).
However, with respect to calendar and human constraints, and taking into account the availability of software components, both in the meteorological community and commercial, some parts of the project stopped early to deal with OOA.

2. OVERVIEW OF SUB-SYSTEMS

During the analysis phase, the project has been split into several sub-systems. Some of these were derived from data flow analysis:

- Preprocessing of observations;
- Observations and point-forecast (#fields) management (BDM);
- Numerical Models;
- NWP fields management (BDAP);
- Product-making and management (FPE/BDPE);
- Transfer agent: product and file broadcast;
- Random access interface (IAA): BDM/BDAP/BDPE/... data service.

The other sub-systems were related to control at different levels:

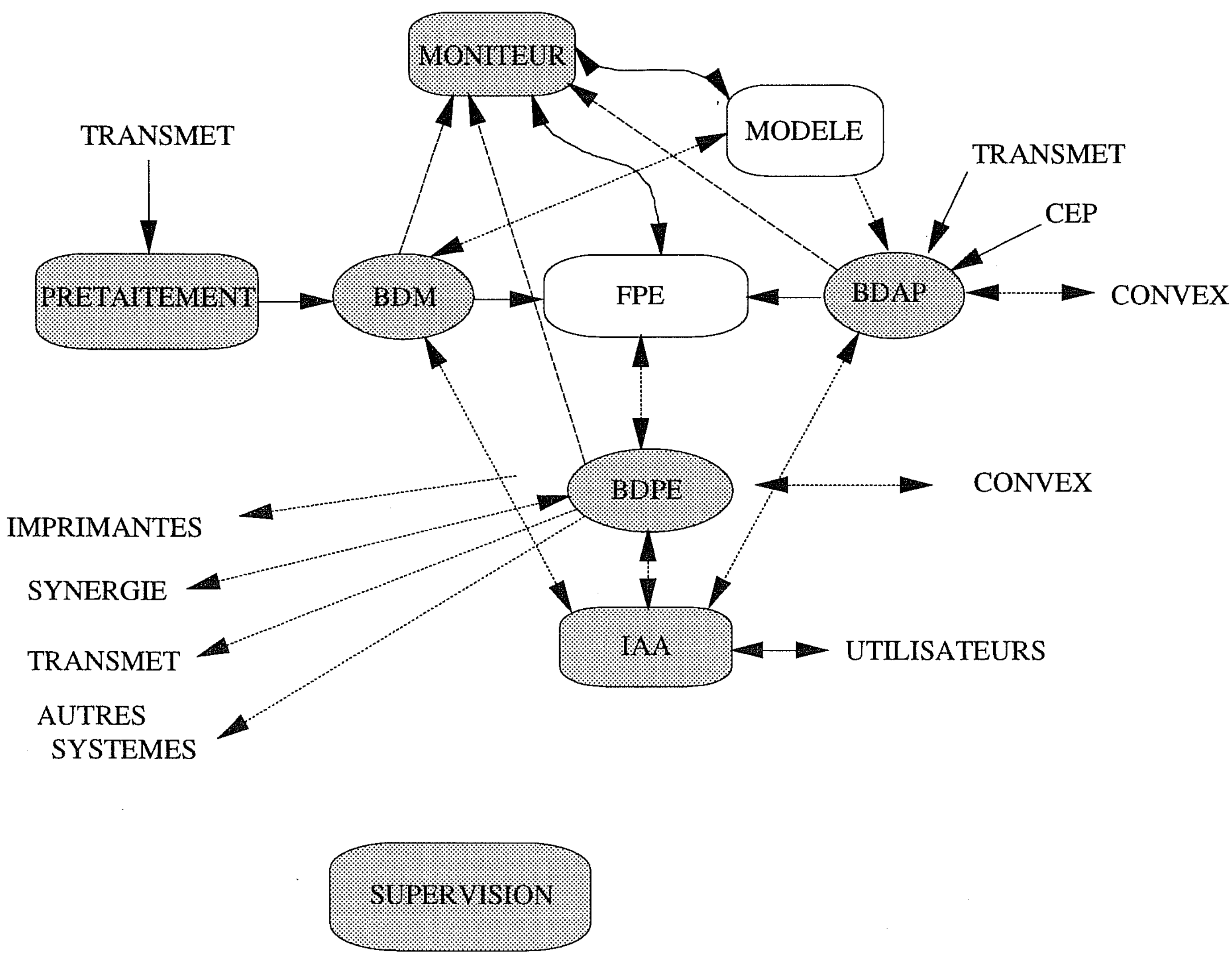
- Métronome: monitor scheduler of tasks;
- Administration;
- Supervision.

2.1 **Preprocessing of observations**

As the first input processor, this sub-system had to be realised rather quickly, because without it most other sub-systems could not be tested nor validated. Due to the wide range of specific codes to handle, the work to be done was considered to be huge. So we used as a basis the ECMWF preprocessing, which was running on VAX/VMS. We ported it into UNIX, interfaces were adapted, streams for observations types were added or upgraded, statistics have been implemented.
Special attention was paid for allowing the feeding of several systems (DIAPASON-like) with pre-processed data.

2.2 **BDM**

Observations management was also a very important part of the project. From the beginning we were convinced of the interest of a commercial Data Base Management System, as a follow-up from previous ECMWF workshops and because of the wide variety and the rather small granularity of conventional observations. But as for other centers, things were not so clear for satellite data.



UTILISATION DE L'AGENT DE TRANSFERT

AVIS DE RESSOURCES AU MONITEUR

Paying attention to the NEONS software form the US Naval Research Laboratory, we took contact and were allowed to use it. NEONS was designed to manage all meteorological data, not only observations, and it was felt as a major tool that might also be used for workstations. At this time, NEONS was available on top of EMPRESS Relational Data Base Management System. Both for technical and strategic issues, it was decided to adapt it to ORACLE RDBMS.

Local modifications were made on contents of primary data records, using both true BUFR for single observations, together with a subset of extracted parameters, this last feature allowing for extraction tools with selection criteria on values of these extracted parameters.

To be able to monitor some productions such as objective analysis start, counters have been set up to send "resources" to Métronome sub-system.
Extra developments have been performed to manage:

- "point forecast", typically forecast at fixed observation points, which may be seen in data model words as (observations+range+model_identification);
- objective analysis quality control data;
- archival (and retrieve) of data.

Source code in C, interfaces for C and FORTRAN programs.

2.3 Numerical Models

This part was already running under UNIX on a CRAY computer. So the work performed was mainly to adapt job submission and control using NQE (CRAY version of NQS), input/output (obs./fields, analysis QC), and to set up signals for Métronome. Data input has moved from CRAY hyperchannel protocol to TCP/IP (ftp). Data output is performed mainly through the Transfer Agent, with results stored in BDAP, BDM and BDPE.

2.4 BDAP

This subsystem deals with storage and service of NWP fields. For use out of production, it offers a transparent access to fields archived on CONVEX/Unitree.
It is based on NEONS though top level interfaces have been rewritten. Fields handled are latitude-longitude ones, stored in GRIB format, vertical levels are isobaric ones and fixed-height above orography.
Extraction functions enable access to sub-grids with possible interpolation, to vertical and/or temporal blocks, linear combinations, and vertical profiles.

Formatting of input data is performed within Numerical Models subsystem for internal models; for external ones there are two preprocessing steps, one specific for ECMWF models, and one for models received through the GTS.
Fields archival itself is performed as an extraction in FPE subsystem.
Feeding of BDAP sends signals ("resources") to Métronome, with a coarse granularity (model/grid/hour/range) for METEO FRANCE and ECMWF models, at field level for GTS models.
Source code in C, interfaces for C and FORTRAN programs.

2.5 FPE and BDPE

FPE is the subsystem where products are made. It may be seen as a collection of tasks, each task leading generally to a single product. A typical task consist of a small shell-script, invoking a program that extract data from main data bases and perform a result, often in a graphical form (PostScript).
Tasks are submitted through, and perform some dialog, with Métronome monitor scheduler.

BDPE is the name given to the data base of products. Each product is described by a product identification, a range, and a date.
BDPE is one of DIAPASON parts where the object-oriented approach has been widely used, though ORACLE RDBMS is used as a service layer.
A product archival/retrieval service is handled, with data residing on CONVEX/Unitree...
Source code in C++ and C, interfaces for C and FORTRAN programs.

2.6 Transfer Agent

Primarily designed as the product broadcast part of FPE/BDPE, its use has been extended for file output from Numerical Models, and also in external models management. Though it may be seen more as a service layer than a subsystem, it has an important role.
Basically, it is used within DIAPASON to transfer files with ftp and to send signals to clients through rpc.
But for external systems, it may be used to send files through ftp (without signal). Files are spooled, and it handles automatic retries in case of failures.
Source code in C.

2.7 IAA

Though DIAPASON is by definition a production system, there was the need for an interactive access to basic data (BDM/BDAP/BDPE), including archived data, for requests outside DIAPASON and/or its production kernel. This has led to IAA, french acronym for Random Access Interface, service which has been designed also for access to METEO FRANCE Climatological Database.
This subsystem deals externally with ftp and X25 requests (text file commands); file sending may be performed through ftp and z_modem. Clients have associated accounts, allowing statistics and possible facturation.
Users of IAA are both external and internal clients, for interactive data access and/or broadcast backup. IAA terminals are being developed, but the very first available terminal for internal users (including overseas services: Martinique, Guadeloupe, French Guyana...) is MetView from ECMWF, which has been adapted to perform extraction of BDM/BDAP data and visualisation.
IAA scheme has been proposed as a possible model for WMO Distributed Data Bases, in the 1995 meeting of the Working Group of Data Management.
Source code in C.

2.8 Métronome

It is the name given to the monitor scheduler built for DIAPASON. It may be seen as the system heart. It starts tasks, both for FPE and Numerical Models, with criteria which are:

- resources sent by data availability in BDM, BDAP, and BDPE;
- time (UTC or local);
- FPE task signals (end with associated status).

Conjunction of criteria is allowed.

In all cases (but the fully dynamic one: alert model) resources have an associated time, which give the time reference, enabling the pre-setting of tasks. There are also two timeout notions: one for giving up, the other one may be used to check in the configuration if the production may be sent "as is" with a sub-set of (available) associated resources, or if it has an alternative production with other resources and other timeout conditions, while the standard production may be maintained or not. Typical applications of these last two features are:

- graphical presentation of ECMWF Ensemble Prediction cluster fields, where the number of elementary maps vary from day to day with the number of clusters;

- aeronautical charts production, where the backup may be to use data from another grid/origin or older.

As suggested in the last example, several levels of alternative production may be specified.

Configuration of scheduling is held through a RDBMS. This is necessary because of the big number of tasks (several thousands). Replication tools available with RDBMS eases setting up of new tasks.
Special care has been given to be able to run in parallel a second monitor, on a separate machine, that do everything but submit tasks, and allowing fast switch-over.

Code development has been made mainly in C++, with some layers in C.

DCE (Distributed Computing Environment) from OSF (Open Software Foundation) has been used for communication/networking but also for architecture: DCE threads are being used. ORACLE RDBMS has been used both for configuration and for "persistant objects".
This last use of RDBMS, giving an image of monitor scheduler objects, has enabled a clear separation between scheduler itself and the supervision of Métronome, the visualisation part being performed through ORACLE GRAPHICS tool. All actions related directly to task execution may be performed through Métronome own supervision.

2.9 **Administration**

This part aims at giving centralized tools for setting up or switching off tasks with related configuration at Metronome/BDPE/Transfer Agent level in a consistent way, and with few manipulations.
The development has been made mainly with ORACLE tools (such as FORMS4).

2.10 **Supervision**

This sub-system allows operators to watch and monitor sub-systems, but also network and machines, to change status of machines (operational/backup/ maintenance), to stop and restart whole subsystems.
It has been split into two parts, leading to separate terminals: an applicative part, and a system part. The applicative part relies on HP Operation Center software. This software mainly deals with alarms sent from sub-systems, but the network is also checked, and for use in the system part vital processes activity and file systems are monitored. The system part is mainly built on HP Glance tool.

3. HARDWARE IMPLEMENTATION AND NETWORKING

The DIAPASON system may be split into three sets of distinct computers:

- a production set;
- a test, integration and adminitration set;
- a development set.

The production set consists of 3 pairs of dedicated machines, and a super- computer shared with Research users. Within each pair, one machine acts as the operational one, the other acting as backup. All databases are fed, but no data comes automatically out of the backup part: no product making nor external routing. However, backup sub-systems are those normally adressed by IAA requests.
All machines specific for DIAPASON are Hewlett-Packard computers.

- One pair for Preprocessing/BDM/BDAP/major part of FPE:
  2x T500 (HP800 servers), 4 proc. 90 Mhz, 2 Gb RAM, 72 Gb disk.
- One pair for BDPE/Transfer Agent/IAA (file servers):
  2x H50 (HP800 servers), 1 proc. 96 Mhz, 256 Mb RAM, 14 Gb disk.
- One pair for Métronome/Supervision:
  2x 715/100XC (HP700 workstations), 1 proc. 90 Mhz, 256 Mb RAM, 6 Gb disk.
- One CRAY-c98, 8 vector proc. 245 Mhz, 2 Gb RAM, 4 Gb SSD, 70 Gb disk, for Numerical Weather Prediction (NWP) Models and a small part of FPE.

The test, integration and administration set, which aims to give a mold for testing new versions of both production tasks and sub-systems, and holds task programs source codes and administration databases, consist of a single machine:

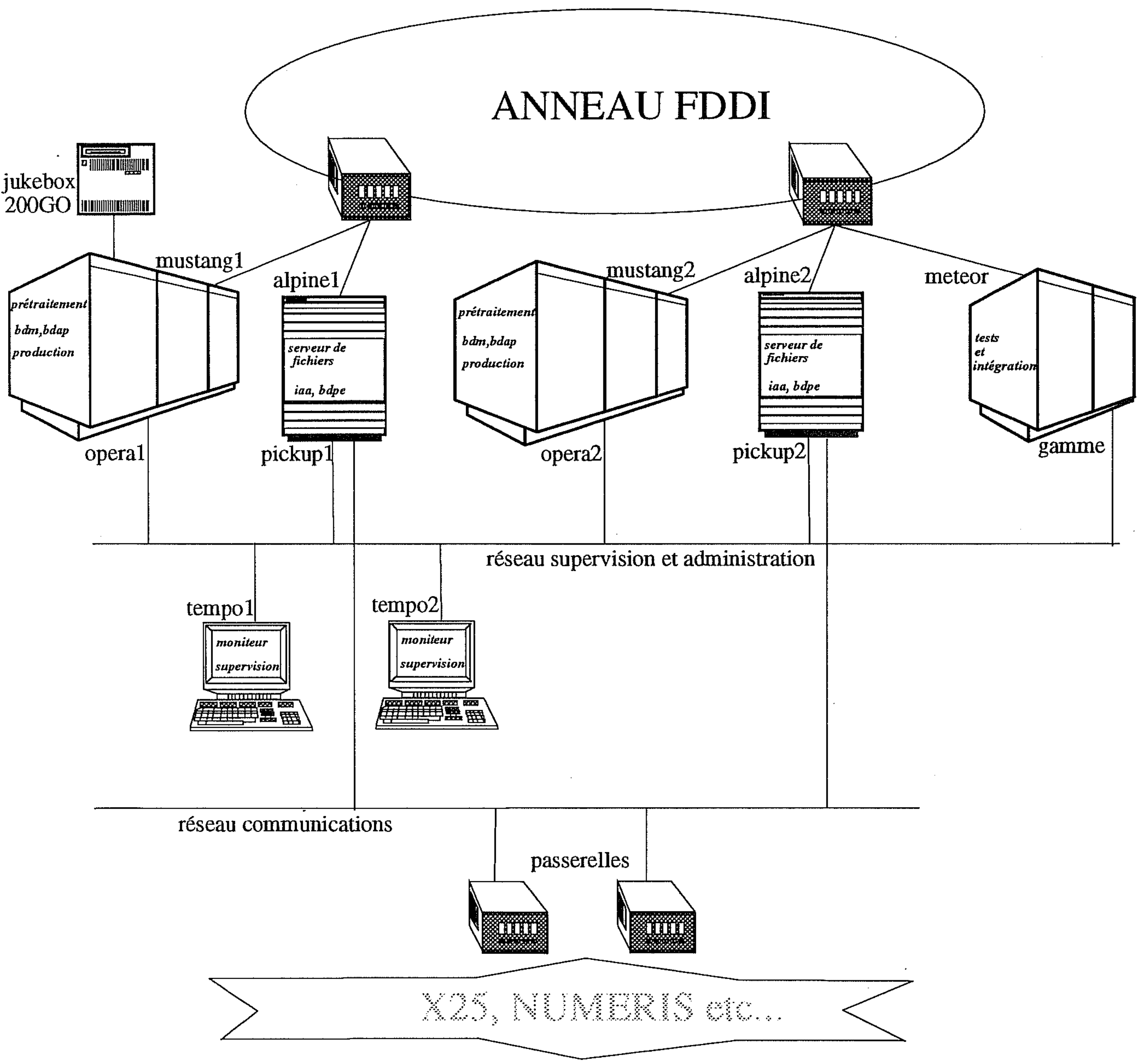
- T500 (HP800 server), 2 proc. 90 Mhz, 1 Gb RAM, 30 Gb disk.

The development set consist of five servers, used both for development both of new tasks and new versions of some sub-systems components:

- 3x H50 + 2x G50 (both HP800), 384 Mb RAM, from 12 to 20 Gb disk.

On these machines, user interfaces to BDM/BDAP/BDPE are the same as those offered in the production set, but access to databases is performed through ORACLE SQL*Net, in read-only and on backup machines.

CONFIGURATION MACHINES DIAPASON



In terms of networking, a dedicated Ethernet network connects the production set (CRAY excepted) and the Test, integration and administration set for supervision and administration.

For data transfer, all machines but Métronome/Supervision use a pre-existing FDDI ring, which is also the path for BDAP/NWP/BDPE archive on CONVEX/Unitree.

Development servers are also connected to other Ethernet networks, in a coherent way with the location of their main users (from X-terminals).

A 200 Gb optical juke-box enables archival of BDM data.

4. OVERVIEW OF "BASIC" SOFTWARE (RE)USED

For production, test, integration and administration:

- preprocessing and decoders from ECMWF;
- NEONS from US Naval Research Laboratory for BDM/BDAP;
- ORACLE RDBMS and related tools for BDM/BDAP/BDPE/Métronome/Adminitration;
- PV-Wave from Visual Numerics and FrameMaker for graphical production;
- NQE (CRAY) and Task Broker (HP) for job submission and control;
- Operation Center and Glance from HP for supervision;
- DCE from OSF for Métronome.

For system development and/or users (teams developing new products):

- Teamwork from Cadre Technology for Computer-Aided Software (CASE) tool;
- MetView from ECMWF for BDM/BDAP extraction and visualisation;
- Applix tools for desktop applications;
- ClearCase for management of NWP models source code configurations;
- other packages or libraries: IMSL from Visual Numerics, MAGICS from ECMWF...

5. STATUS AND CONCLUSIONS

The DIAPASON system has replaced the previous data processing system (CDC/NOS) at the end of august, 1995. The move of operational production to DIAPASON has been made step by step during the last monthes.

Some extra work is still to be done, for instance tunings and some optimisations. Project team is progressively "giving keys" to administration team.

Though not yet closed, it was a major project for METEO FRANCE, 3 year and a half long, with about 50 men-year of human resources (time for port of specific productions not included) and a whole budget of nearly 30 millions FF. It has led indirectly to a new organisation of central computing services.

About 2300 tasks are configured, leading each day to about 5000 jobs, and 1650 graphical products.

Clearly, DIAPASON has been a major upgrade for METEO FRANCE production. Improvements may be split in two parts:

Quantitative part:

- More basic data are available, for use by objective analysis and products:
    - Extra observation types: ACARS, TOVS120, ERS1, profilers, radiances...
    - NWP models output: more parameters, levels and grids.
- More products are already being made:
    - Graphical products for the Central Forecast Service;
    - Models fields for SYNERGIE workstations, +soon in new channels of RETIM satellite broadcast.

Qualitative part:

- Enhanced reliability:
    - Dedicated machines organised in pairs;
    - Clean integration of previous out-of-system productions;
    - Use or reuse of solid external software;
    - Better regularity after tunings/optimisations;
    - Evolution now possible, both at hardware and software level.
- Products have been revisited.
- Homogeneous with message switch (TRANSMET) and forecasters workstations (SYNERGIE).
- No more preventive maintenance.