EE

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

8u 9401

CERN - SL DIVISION

9

CERN LIBRARIES, GENEVA

CERN SL/93-42 (CO)

CM-P00061194

LEP Accelerator Logging System using on-line Database

R. Billen, F. Bordry, C. Delamare, E. Hatziangeli, F. Roeber
CERN, Geneva, Switzerland

ABSTRACT

The performance and efficiency of LEP depend on a multitude of factors, including particle beam characteristics, physics parameters, hardware settings, and environmental conditions. Often, these factors interact in unexpected ways and affect the machine performance.

In January 1992 a project was started to create a unique logging system using an on-line database. One year's worth of data was to be kept on-line, this was estimated to be 8 GB. The systems concerned are of different nature (particle beam profile, power converter current, meteorological data, magnet temperature, lep-mode,...), different frequency (from seconds to several hours) and different size (from 10 MB to 1 GB per year). Major performance criteria included rapid logging of data for useful real-time monitoring of compound measurements, and rapid retrieval and correlation of large amounts of data for efficient off-line analysis.

For the database design the NIAM methodology was used as well as some interesting techniques such as tagging the rows with timeslots instead of timestamps and row packing for storage minimisation. A great attention was paid to the definition of the space allocation for tables and indices since this is a determinant factor for the efficiency of a large database (the largest table has more than three million rows). A complex structure of servers and clients takes care of data gathering, data logging and management of all real time measurement and logging requests. Several tools have been developed to make the data correlation transparent to non-database experts.

ICALEPCS'93

International Conference on Accelerator and large Experimental Physics Controls Systems
Berlin, Germany, October 18-22, 1993

LEP ACCELERATOR LOGGING SYSTEM USING ON-LINE DATABASE

R. Billen, F. Bordry, C. Delamare, E. Hatziangeli, F. Roeber
CERN - European Laboratory for Particle Physics

Abstract

The performance and efficiency of LEP depend on a multitude of factors, including particle beam characteristics, physics parameters, hardware settings, and environmental conditions. Often, these factors interact in unexpected ways and affect the machine performance.

In January 1992 a project was started to create a unique logging system using an on-line database. One year's worth of data was to be kept on-line, this was estimated to be 8 GB. The systems concerned are of different nature (particle beam profile, power converter current, meteorological data, magnet temperature, lep-mode,...), different frequency (from seconds to several hours) and different size (from 10 MB to 1 GB per year). Major performance criteria included rapid logging of data for useful real-time monitoring of compound measurements, and rapid retrieval and correlation of large amounts of data for efficient off-line analysis.

For the database design the NIAM methodology was used as well as some interesting techniques such as tagging the rows with timeslots instead of timestamps and row packing for storage minimisation. A great attention was paid to the definition of the space allocation for tables and indices since this is a determinant factor for the efficiency of a large database (the largest table has more than three million rows). A complex structure of servers and clients takes care of data gathering, data logging and management of all real time measurement and logging requests. Several tools have been developed to make the data correlation transparent to non-database experts.

1. Introduction

The e+e- collider LEP is used to investigate the Z^0 particle and to measure its energy peak and width. The mass of the Z^0 boson is determined by combining the data from the four LEP experiments. Thus the aim of LEP operation is to get a high integrated luminosity. The performance and efficiency of LEP depend on a multitude of factors, including particle beam characteristics, physics parameters, hardware settings, and environmental conditions. Often, these factors interact in unexpected ways and affect the machine performance. To optimise the accelerator performance, it is necessary to correlate all these parameters. Furthermore the dominant error on the Z^0 mass arises from uncertainties in the calibration of the energy of the beams in LEP. The absolute energy is determined by resonant depolarisation of transversely polarised beams [1]. This measurement is a time consuming operation (several hours, once or twice a week) and is preventing, for now, the data taking by the experiments. To provide the energy calibration of all LEP data, corrections to the measured field display energy in a reference magnet are applied based on magnet temperatures around the ring, RF voltage distribution, etc.[2].

During the first two years of LEP operation, there were many parameters logged (beam currents, luminosity, energy,...) but each system was different and incompatible (structure and medium). Furthermore, there was a huge amount of data duplication as well as missing data. In January 1992 a project was started to create a unique logging system [3]. The first stage of the project was to define a generic structure for the data and to review all the requirements for all the heterogeneous systems. To manage and correlate large amounts of data, a central database was essential, in which all data concerning the different phases of LEP running could be stored and shared. Considerable experience with the Relational Database Management System (RDBMS) ORACLE, which runs on most of CERN's computing platforms has been gained during the construction and the operation of LEP[4].

2. Data description

The variety of data is large due to the heterogeneous character of the data systems. To give a few examples of the data of interest: Intensities of each lepton bunch; Transversal profile of each lepton beam; Transversal position of closed orbit (500 pick-ups); Power converter currents and tolerances (800 power converters); Voltage, power, frequency of accelerating RF-units; Environmental parameters such as temperature, pressure, humidity; Experimental key parameters such as luminosity, background rate; ...

Also the rate of the data taking varies between seconds (beam intensities) and hours (magnet temperatures) as well as asynchronous data (events). Consequently the yearly amount of data to be logged varies drastically per system (from 10 MB to 1 GB per year). One year's worth of data was estimated to be around 8 GB.

An important notion is the lep-mode, i.e. the actual (dynamic) state of the LEP machine. During normal operations of LEP for physics, a sequence of specific lep-modes is followed:

- 1) *set-up* preparation of all equipment for 20 GeV lepton beams

- 2) *filling* injection and accumulation of leptons at 20 GeV
- 3) *acceleration* ramping of all equipment to 45 GeV beams
- 4) *adjust* colliding and tuning of the beams
- 5) *physics* stable conditions at 45 GeV for experimental data taking

For periods without beam lep-modes such as *shutdown* or *access* can be in effect. Clearly the data gathering rates are dependent on the lep-mode. It is possible to change the logging frequency for each data system for each lep-mode. Of course, this implies that each logging process needs to be aware of the actual lep-mode (cf. 4.3. database monitor).

3. Database design

3.1 Design Method and Tool

For the database design the NIAM (Natural Information Analysis Method) was used [5]. In a NIAM schema one easily expresses objects, facts between objects, constraints on facts and between objects. This data analysing method was preferred above the ERD (Entity Relation Diagram) method because of the binary relationship approach, completeness and clarity of the method. Moreover, the RIDL* tool [6] allows a rapid and efficient creation of the database tables starting from a NIAM schema. The schema is easily introduced in the graphical interface, after which a completeness analysis and mapping onto ORACLE tables is done in an automatic manner.

3.2 Timeslot data tagging

For each data system the notion of timestamp as well as timeslot have been introduced at the design level. It is clear that for a given measurement a unique timestamp is essential. An equivalent attribute called timeslot has been added, being a unique integer. For repeating data (cf. 3.3 row packing) only the timeslot tags a set of measurements, while the association between timeslot and timestamp is made in a separate table. The timeslot approach minimises the data volumes but has a major impact on data correlation, especially when searching for specific (non-occurring) timestamps [7].

3.3 Row packing

For large data systems that have simultaneous measurements (e.g. 260 currents of horizontal corrector power converters) a row packing technique has been introduced (13 groups of 20 power converters), rather than to have one table row per converter.

One static table makes the association between one specific equipment (e.g. a power converter) and its group and column position. Each row is tagged by a timeslot only. A view decodes this underlying two-dimensional structure and makes it transparent to the end-user. As a result the data volumes are severely reduced. For the power converter example (800 power converters), the storage is: 1 GB for the data and 30 MB for the indexes with row packing techniques (27 groups of 30 power converters) instead of 2 GB for the data and 1 GB for the indexes.

3.4 Tablespace definition

A separate tablespace has been foreseen for each data system. Data volumes have been calculated on the basis of one year's worth of data taking at a desired maximum rate. Tables have been created with well defined storage parameters to avoid table fragmentation. These tables have been filled with incrementing timeslots and dummy timestamps and dummy data. The logging processes only perform updating of rows. An automatic roll-around is implicit since the data for timeslot 1 is overwritten when the maximum timeslot has been reached. Indexes have been created on timeslot and timestamp columns, which is absolutely vital for fast data access. However the cost in storage space is quite high: the index volume can be as large as the data volume for a simple system where no row-packing is applied.

4. Data Gathering system

Figure 1 illustrates the measurement system, the logging system and the database monitor server (dbm) [8]. The measurement system comprises the measurement server (msrv) and the measurement black boxes (mbbs). The logging system comprises the logging server (lsrv) and the logging black boxes (lbbs).

The aim of the measurement system is to have a single process to acquire measurement data for any given hardware system, and to provide the data to one or many applications programs (hereafter measurement clients).

The logging system is a client of the measurement system. The logging black boxes get measurement data from the measurement black boxes or from the measurement database and they log data in the logging database.

The database monitor is a process which monitors the LEP control database. It is able to inform an application program when the contents of any database table have changed, e.g. when the database table which contains the LEP mode information is changed.

TCP/IP sockets are used for interprocess communication between the measurement server and the measurement black boxes, between the logging black boxes and the measurement black boxes and between the database monitor and its clients.

4.1 Measurement system

The measurement server is a single process which acts as a measurement multiplexer between measurement black boxes which acquire measurement data from the hardware and many measurement clients which want measurement data. It waits on a designated TCP/IP port for connections from measurement clients. Once a client has asked for a measurement, the measurement server starts off the appropriate measurement black box (if it is not already running), which performs this measurement. The measurement server passes then transparently a connection to the measurement black box and it closes its own client connection. The measurement server monitors the state of all the measurement black boxes and guarantees their presence.

In the context of the measurement system, any program which would like to have measurement data from the measurement system is a measurement client. A measurement client is able to specify : whether it wants data directly from the measurement black box or whether the data should be placed in a database table in the Measurement Database, how recent the measurement data should be, and if it wants repeated measurements, the frequency at which a repeat measurement request should be performed.

The measurement black box performs single or periodic measurements. There should be one measurement black box for each hardware system that needs to be measured. A measurement black box may have more than one client and it sends back measurement data to its client at the requested frequency rate. The measurement black box has a standard layout and the authors of the black boxes should only supply the actual measurement routine, which does the specific equipment access.

4.2 Logging system

The logging system comprises the logging server and the logging black boxes.

A logging black box is a client of the measurement system. It gets measurement data from a measurement black box and it logs it in the Logging Database. A logging black box process can log continuously and conditionally (hereafter *condition to log*), e.g. the luminosity logging black box should log whenever the LEP mode is *adjust* or *physics* and it should not log at any other time.

The logging server is a single process. It starts any logging black boxes whose *condition to log* is satisfied. It monitors the state of all logging black boxes and guarantees their presence during the period when their *condition to log* is true. When a *condition to log* becomes false, the logging server signals the corresponding black box to stop its execution. When a *condition to log* becomes true, the logging server signals the corresponding black box to resume its execution. The logging black boxes handle properly the signals sent by the logging server.

4.3 Database monitor

The database monitor server is a single process which waits on a designated TCP/IP port for a connection from any application program. This server accepts connections from clients who are interested in finding out when Oracle tables change in some way. When it detects a change in a table, it informs the appropriate clients. The measurement server and the logging server are clients of the database monitor server, as well as most of the logging black boxes.

4.4 Surveillance and consistency of the data logging

When LEP is coasting off-peak (i.e. the energy given to the particles is either below or above the Z^0 mass) the importance of the logging system is absolutely crucial for the validity of the gathered experimental data. Parameters such as RF-voltage distribution and magnet temperatures are vital to be able to apply energy corrections to the experimental results. At the end of each coast the LEP experiments analyse the availability of the logged data and decide whether the next coast is allowed at an off-peak energy or not.

The surrounding software structure of Measurement and Logging Server supervising (baby-sitting) the data gathering and logging processes already fulfils a first level of security. However no direct information is available on a malfunctioning of one of the elements, nor on storing of corrupted data. Therefore a surveillance program is installed to monitor the availability as well as the consistency of the logged data. Every data system (RF, Power Converters,...) has its proper set of parameters concerning the necessary freshness and value range of the logged data with respect to the lep-modes. The allowed tolerances on the data are very tight to allow a rapid identification of a potential logging problem. The surveillance program reports freshness and range anomalies to the LEP Alarm System. An absence of the surveillance program is also detected and reported by the LEP Alarm System. The operations crew have instructions to diagnose and solve a logging problem.

5. Database implementation and administration

5.1 Hardware

A dedicated machine was purchased to install the LEP logging database. A SUN workstation 630 MP with two CPUs (20 MIPS, 64 MB of memory.) was chosen. Due to the large amount of data to log and to keep on-line, a total of 14.3 GB of disks is used (Five 1.3 GB disks for database software and data, four 1.3 GB disks for database backup).

5.2 Software

The Database management system is Oracle Version 6.0.36.4.1 and the Operating systems SunOS 4.1.2. The plan is to change for the end of the year to Oracle Version 7 and to Solaris operating system.

5.3 Database administration

The Oracle database is running in the so-called "archivelog mode", which means that all updates to the data are saved so that they can be reconstructed in the event of media failure. These updates are stored in "off-line redo log files" on disk, with a unique log sequence number for each file.

Backups have to be frequently performed, because the older the backup, the more off-line redo logs are required to recover. In the course of normal database operation, the redo log files record all transactions following a backup until the next backup is taken. "Backup" means "off-line backup", i.e. performed while the database is shutdown. Since the LEP accelerator is typically stopped at least once a week, either for scheduled maintenance or unscheduled problem, the backup procedure is then run manually during this time. A mail message is automatically sent to a list of users to inform them about the database shutdown. About 60 minutes are required to backup the complete database (5GB in volume) from disk to disk. In their turn, the four backup disks are backed up via the IBM tool DFDSM (Data Facility Distributed Storage Manager) which transfers the data onto an IBM/VM robot cartridge system.

An exabyte stacker directly accessing from the Sun station (10 tapes of 5GB each) is going to be exploited in order to produce extra safe copies of the four backup disks. This tool is also used to backup the off-line redo log files which are vital prerequisites in case of a recovery. Automatic procedures are run to perform the backup of the off-line redo logs. The disk containing the off-line redo logs is also subject to automatic surveillance, because an overflow of this disk could cause the entire database to hang.

5.4 Experience and problems

A high stability of the database system has been observed up to now, together with a good performance and quick response time to the users. The recovery exercise was run successfully once in the 14 months of production. A hardware problem on a disk was probably at the origin of data corruption. An almost automatic operation of the database has been achieved, based on a collection of automated procedures. The database administrator intervenes when updates to the internal data architecture or recovery exercises are needed.

6. Data retrieval

The LEP logging project was developed to provide a user-friendly tool to a large spread of users. The main users are the accelerator physicists (optimisation, machine development results, beam-beam,...), the accelerator operators (reproducibility, optimisation, statistics,...), the experiment physicists (energy calibration, luminosity and background studies...), various people of equipment groups (beam instrumentation, RF, power converters, separators,...). For all these end-users, the access to the logged data must be transparent (without any knowledge of Oracle nor SQL) and fast (the time response has to be in the order of one minute for retrieving data of a typical 15 hours fill). These constraints apply to correlations between heterogeneous systems (different logging frequency and data).

The timeslot/timestamp concept was chosen to permit efficient correlations when huge tables are handled. To correlate two different systems, the first idea was to realise the correlation using the ORACLE kernel. The following SQL statement illustrates the correlation between beam intensities and luminosity of one experiment (in this example ALEPH) :

```
select to_char(bct.timestamp,'dd-mm-yy hh24:mi:ss') time
       ,Ie , Ip, Idc, aleph1.luminosity aleph_lumi
from   bct_data bct, aleph_luminosity aleph1
       ,aleph_luminosity_timestamp aleph1, aleph_luminosity_timestamp aleph2
where  bct.timestamp > (select min(timestamp) from fd_data
                        where lep_mode = 'physics' and fill_number = &Fill_Number * 100)
and    bct.timestamp < (select min(timestamp) from fd_data
                        where fill_number = (&Fill_Number + 1) * 100)
and    aleph1.timeslot = aleph2.timeslot - 1
and    aleph1.timestamp <= bct.timestamp
and    aleph2.timestamp > bct.timestamp
and    aleph1.timeslot = aleph1.timeslot
and    abs(86400*(aleph1.timestamp - bct.timestamp)) < 600 ;
```

The time response obtained was in the order of one hour for a long fill. The correlation of ten systems took more than one day ! Several performance tunings reduced these times but it was very far from the target time. Finally it was decided to do the correlation process outside of Oracle using a C program filter. The use of the UNION operator and the filter program provided a reduction of the time

response from one hour to 10 seconds. The biggest data correlation (10-15 systems) is done in less than 2 minutes with this technique.

Interfaces for end-users were implemented on an IBM/VM mainframe and on UNIX workstations, by means of hidden predefined views and SQL statements. The outputs of these interfaces are ASCII files; the users can plot their data using their favourite graphical tools (Excel, Paw, ...). A uniform graphical user interface is under development.

7. Conclusion

Prototype versions of the measurement and logging systems have been logging a subset of the required data in 1992, and the performances confirmed the validity of the database design. Since the 1993 LEP start-up in May, the logging system is operational and gradually more and more parameters are logged. The logging database is an important tool for LEP optimisation and development. The number of users is growing every day. During peak activities, more than fifty users and twenty logging black boxes are connected to the Oracle database without performance problems.

Furthermore, in the context of the energy scanning for a better determination of the Z^0 mass and width the logging database became a crucial system.

8. Acknowledgements

The authors would like to express their gratitude to S. Santiago for his innovating ideas and his assistance, C. Ellis (ORACLE UK) for his original ideas and D. Ensor (ORACLE UK) for performance tuning.

The authors are grateful to A. Silverman, I. Reguerro for hardware support and to N. Segura-Chinchilla for database support.

9. References

- [1] L. Arnaudon and al., "Measurement of LEP beam energy by resonant depolarisation", Phys. Lett. B 284 (1992) 431
- [2] LEP Energy Working Group, "The Energy Calibration of LEP in 1992", CERN-SL/93-21(DI), April 93
- [3] J. Poole, "Proceedings of the second workshop on LEP performance", CERN-SL/92-29(DI)
- [4] R. Bailey, "Development of the LEP high level control system using Oracle as on-line database", ICALEPCS'93 - Berlin, October 1993
- [5] J.J.V.R. Wintraecken, "The NIAM Information Analysis Method - Theory and Practice" (Kluwer Academic Publishers, 1985)
- [6] RIDL* (V1.2) manual set, IntelliBase nv/sa, Plantin en Moretuslei 220 bus 3, 2018 Antwerpen, Belgium (1991)
- [7] R. Billen & F. Bordry, "LEP Logging : Repeating data description Timeslot and row packing concepts" CERN Note(1992)
- [8] E. Hatziangeli, F. Roeber, "A description of the LEP measurement and logging system and the application Interface libraries", CERN SL/93-38(CO)

Fig 1. The Measurement and Logging System