Distributed Objects in Action - First Experience with the CLEO III Data Acquisition and Control System

H. Schwarthoff[4], *R. Baker*[1], *C. Bebek*[1], *V. Frolov*[5], *C. Gwon*[4], *K. Honscheid*[4], *D. Hufnagel*[4], *D. Kreinick*[1], *E. Lipeles*[2], *T. Pedlar*[4], *C. Strohman*[1], *J. Thayer*[4], *A. Weinstein*[2], *A. Wolf*[3]

[1] Cornell University, USA,
[2] California Institute of Technology, USA,
[3] Syracuse University, USA
[4] The Ohio State University, USA
[5] University of Minnesota, USA

Abstract

The new CLEO III experiment at the Cornell electron positron storage ring CESR includes 4 layer silicon vertex detector and a novel ring image Cherenkov detector. Driftchamber, electromagnetic calorimeter and muon chambers complete the detector assembly. Overall more than half a million electronic channels have to be read out and monitored. Detector configuration, data quality and component monitoring, run control are only some of the Slow Control tasks that have to be performed.

By the time of CHEP 2000 the initial commissioning run will be completed. In our presentation we will focus on our experience with a large distributed system with special emphasis on our success - and problems - with industry standards such as CORBA (inter-platform communication), Java (remote access via Web browser) and Objectivity (event and constants database). We will discuss our design and highlight some of the special or unique features of the CLEO III monitoring system.

Keywords: DAQ system, CORBA, DBMS

1 Introduction

The CLEO detector at the 10.6 GeV e^+e^- collider CESR at the Cornell University in Ithaca, New York is about to complete a major upgrade. The CLEO III detector ([1]) includes a new Si vertex detector and a particle identification system based on *Ring Imaging Cherenkov* (RICH) detectors. In parallel, modifications to CESR will result in an increase of the peak luminosity to $3 \times 10^{33} \text{cm}^{-2}\text{s}^{-1}$.

In order to handle the increased amount of data to be processed and written to mass storage a new trigger, readout and control system has been designed and implemented. The data rates after the first trigger level are expected to be around 40 MB/s.

2 The new CLEO III detector

The CLEO III detector has been read out with the newly installed RICH detector and Tracking Drift Chamber, along with the existing Cesium Iodide Crystal Calorimeter. The new Silicon Vertex detector is being installed in February 2000, and the existing Muon Chambers will resume operation.

All detector data is read out through Motorola VME PowerPC CPUs and routed to a central *Event Builder* through standard 100baseT Ethernet connections. The transfer rate has been measured to be 8–9 MBytes/s. Data storage is provided from there through 36GB AIT tape cartridge drives.

The readout and control software has been developed using the new CLEO III C++ coding environment. Event data is transported via TCP/IP socket connections, while the Slow Control functionality is provided through *CORBA* middleware ([2]) messaging. This enables us to run a homogeneous network of components with nodes running a variety of operating systems (Unix, Windows NT, and VxWorks). The *VisiBroker* CORBA implementation ([3]) has been used. Table I illustrates the three platforms in operation. All persistent data that is not event related is stored through a customized set of object databases, implemented with the *Objectivity* Object Database Management System ([5]). The Java 2 platform serves as basis for a set of Graphical User Interfaces (GUIs) for operation and monitoring. The GUIs are interacting with the control system via CORBA. The system is characterized by a high degree of parallelization (multithreading and task synchronization through CORBA).

Table I: The CLEO III Online computing platforms and their functions

Hardware	**Operating system**	**Function**
Motorola PowerPC	VxWorks 5.3.1	Data board readout, hardware controlling
Sun Sparc	Unix (Solaris 2.6)	Event building, Central Slow Control, Database, Event display
Intel Pentium	Windows NT 4.0	Hardware Controlling, GUIs

The CLEO III Online computing system is isolated from the rest of the world. Its nodes are installed as part of a ”private” subnet and bridged to the ”public” side through a single router. This router is specifically configured to prevent unwanted traffic from entering the ”private” subnet, so that the full Ethernet bandwidth is available for the detector operation. This setup ensures the high degree of self-sufficiency and network performance that is needed to operate CLEO III as a stable, reliable system.

2.1 Detector control

A set of Slow Control components provides the CORBA services used for detector control. The following list gives a simplified overview, illustrating the procedure in a data taking run:

- The *Configuration Manager* surveys the state of all involved software components and ensures the integrity of the system.
- The *Alarm Manager* is the central switchboard for alarm messages and is connected to the Alarm Database.
- The *Run Controller* carries out run state transitions like *Begin Run* or *End Run* and delivers them to all components that take part in the run, initiated from the *Run Control GUI*. It records the basic run information in the Run Database.
- The *Information Manager* API allows all components to publish vital information through CORBA, so that it can be processed or displayed by other components or GUIs.
- The *Run Statistics Collector* collects data that is relevant for later analysis from all existing *Information Managers* and stores it in the Run Statistics Database.
- A set of *Constants Database Servers* allows storage and retrieval of detector constants[1]. Constants are downloaded/verified at run state transitions, initiated by the *Configuration Manager*. An additional database server provides the program executables for most components.
- A dedicated data monitor samples the event stream and performs online diagnostics and monitoring tasks. Those tasks are executed in an online version of the regular CLEO III

[1] The CLEO III constants database system is presented elsewhere in this conference.

data reconstruction and analysis framework[2], so that standard analysis algorithms can be applied here.

2.2 Preliminary system configuration

The nearly complete system consists of 16 VME PowerPC CPUs that are reading out the detector data from the following configuration:

- 8 VME crates, connected to the 230000 readout channels of the RICH detector
- 4 FASTBUS crates, connected to the 7800 readout channels of the Crystal Calorimeter, and routed to 4 VME PowerPC CPUs via a dedicated interface (*FRITZ*, [4])
- 4 FASTBUS crates, connected to the 10000 readout channels of the Tracking Drift Chamber, also routed to 4 VME PowerPC CPUs through FRITZ.

Seven additional VME PowerPC CPUs perform additional slow control tasks. The central readout and control system is run on three Unix (Solaris) machines that perform event building, data storage, database processing, and a number of central controlling tasks, like Alarm messaging and system configuration. Other hardware controllers (for gas and cooling systems) have been realized on three Windows NT (Intel) nodes. The java based operation and monitoring tasks are performed by running the GUIs on Windows NT machines, and, to a lesser extent, on Unix computers (Sun Solaris).

3 First experiences

After a design and development phase of ca. 3 years first experiences have been gathered. The full system could never be tested in its full size. In particular a large number of VxWorks CPUs operating at the same time poses a challenge to the speed and stability of the assembly. It was therefore necessary to go through a considerable shakedown phase to gain a full understanding of the strong and weak points and obtain a stable and powerful configuration.

3.1 CORBA

The *VisiBroker* CORBA implementation has been available in a special release for the VxWorks real time operating system. In general the interoperation with the other platforms is good. However, some specific issues have to be addressed:

- On the PowerPC platform memory management and task cleanup must be handled carefully, since a VxWorks task (i.e. a thread) is not encapsulated against the rest of the system, like it is the case in Unix or Windows processes. Special consideration is required for initialization and shutdown of the CORBA environment, the core of which had to be integrated into the VxWorks kernel.
- In the event of a PowerPC CPU crash or reboot allocated resources like IP sockets are not guaranteed to be cleaned up properly. This can lead to locking situations that can affect other components on other nodes, e.g. central control processes or the CORBA naming service. Here it is necessary to intelligently manage resources and provide timeout mechanisms for all connected processes to allow them to recover.

3.2 Objectivity Object Database

All implemented Online database processes are part of the Slow Control system. Their presence is managed by the VisiBroker *Object Activation Daemon*, which ensures that requested CORBA

[2] Several topics about the CLEO III Offline reconstruction software are presented elsewhere in this conference.

interface implementations are active. The database storage format is controlled through a rigorous build system, necessary to prevent database corruption. The individual processes have been tailored to perform well when all particle detectors are started. At that time configuration and constants data as well as program code must be downloaded simultaneously for all components. The time needed for this procedure must be minimized by allowing multiple parallel database readers.

3.3 Java

The Java 2 platform is used for all Graphical User Interfaces in connection with the *VisiBroker for Java* implementation. The GUI classes inherit from a generic base class *CLEOGui* that provides some common functionality like access authorization or CORBA connection status. Due to the complexity of the monitoring tasks desired most GUIs are highly multithreaded. The management of all threads within a GUI must be performance optimized and fault tolerant to ensure usability.

4 Future plans

The CLEO III detector readout and control system will be completed by adding 5 new VME crates with one PowerPC board each to read out the 125000 channels of the Silicon Vertex Detector. By that time we will have gathered sufficient expertise to be able to provide a well understood and powerful environment that allows CLEO to exploit the new research opportunities that the phase III upgrade provides. We expect to resume the collection of Physics data in April 2000.

5 Acknowledgements

This work has been supported by the US Department of Energy, the National Science Foundation and the Alexander von Humboldt foundation.

References

1 *The CLEO III detector, design and physics goals.* CLNS Report 94/1277.

2 CORBA (Common Object Request Broker Architecture) is an industry standard for remote method invocation, see `http://www.omg.org`.

3 VisiBroker by Inprise, Version 3.2. References can be found on `http://www.inprise.com/visibroker`. The VxWorks implementation is marketed separately by Highlander Communications (`http://www.highlander.com`.

4 FRITZ, a Fastbus Readout Interface with data Translation and Zero suppression. Designed and built at The Ohio State University.

5 Objectivity object database, by Objectivity Inc., Version 5.2. See `http://www.objectivity.com`.