To be presented at IADBG Workshop on Databases for Accelerators (International Accelerator DataBase Group), Chicago, IL, 11/6-8/95.

CONF-9511152--

BNL-62365

Database Activities at Brookhaven National Laboratory

C. G. Trahern*
Brookhaven National Laboratory, P. O. Box 5000
Upton, New York 11973

Abstract

Brookhaven National Laboratory is a multi-disciplinary lab in the DOE system of research laboratories. Database activities are correspondingly diverse within the restrictions imposed by the dominant relational database paradigm. We discuss database related activities and tools used in RHIC and in the other major projects at BNL.

1 Overview of BNL Database Activities

There are three major technical information systems at Brookhaven National Laboratory which use commercial relational databases as critical elements. One of these is an accelerator project, the Relativistic Heavy Ion Collider and Alternating Gradient Synchrotron (RHIC/AGS). The others are the Protein Data Bank being maintained by the Chemistry department, and a Geographical Information System (GIS) – a Super-fund sponsored environmental monitoring project under development in the Office of Environmental Restoration.

In addition to these technical information systems, commercial databases are being used in other areas at BNL. For example, the lab purchasing, inventory and payroll systems currently exist on a proprietary database system, known as *Image*, provided by Hewlett-Packard and running on a HP 3000 system. There are about 200 simultaneous users connected via PC's. These administrative systems are now under review with the aim of replacing them with other products incorporating standard relational database systems and interface tools. For example changing from the HP 3000 system to a NT based database server with *PowerBuilder* applications may be one upgrade path.

Administrative systems at the National Synchrotron Light Source (NSLS) such as the coordination of user groups and experiments as well as basic personnel data are being maintained using Informix on an HP 9000 700 series machine. Finally, there is a large data migration project underway in the Occupational Medicine Clinic. Medical records from the past years are being transferred from a paper based system to a relational database and will be updated in the future via an automated forms interface probably via *PowerBuilder* or *Oracle*Forms*.

All three of the major relational database vendors, Oracle, Sybase and Informix are represented in these systems – leading to the obvious conclusion that there is no laboratory standard for relational databases at BNL. On the other hand there have been efforts to recommend and provide support for standardized interface tools which are compatible with all these database products. In addition to these tools, there are also specialized interfaces. For example, the GIS uses a proprietary geographical interface called *Arc/Info*[5].

The following multi-vendor database interface and design tools are being used at BNL.

- ReportSmith[1]
- PowerBuilder[2]
- PassPort[3]
- Erdraw and Object-Protocol Model toolset from LBL[4].
- Arc/Info[5]

ReportSmith is a report writing utility that works with most RDBMS's. The output from an SQL query can be formatted as one likes and is primarily directed to paper reports. *PowerBuilder* and *PassPort* are graphical programming environments for building sophisticated forms. *PowerBuilder* is currently available on only on Intel architectures, but will be available on SUN platforms in November '95. It is an object-oriented programming environment and works with most RDBMS's.

*Work performed under the auspices of the U. S. Department of Energy

PassPort is an X11/Unix based interface builder. It is being used at the NSLS to provide a touch-sensitive screen directory of NSLS personnel through Informix. Finally, although not multi-vendor, Oracle[6] products such as *Oracle*Reports*, *Oracle*Forms*, etc. are also being used at BNL, particularly in the medical records upgrade.

2 RHIC

RHIC is using Sybase as its relational database management system. It runs on a two processor SUN Sparcserver 1000. A variety of configuration databases have been constructed to manage technical information such as lattice design, magnetic field measurements, impedance, survey and alignment measurements, front end computer configurations, cable layout and tunnel installation. Many of these databases have been described in previous reports[7][8][9].

At RHIC the basic pattern of the data "life cycle" in these configuration systems is probably very familiar. Data is generated by various programs such as *MAD* (e.g. in the case of the lattice design) or in a test facility (e.g. in the case of magnet field measurements). Tabular data structures are defined using tools such as *Erdraw* or *OPM*, including whatever referential integrity constraints that may be appropriate. The data are then organized into ASCII files according to the database table structures or into binary C/C++ data structures that correspond to many tables. The data are then entered via a Sybase supplied file transfer utility in the case of ASCII files or in the case of binary data structures via locally written C/C++ programs that use Sybase supplied Open Client C libraries. The database tables are updated on a regular basis depending on the rate of change of the source data. At this point C/C++ application programs are written to generate reports or to derive other data that will in turn flow into other databases/tables. In all these areas of data management, very little time critical transaction processing is involved but reasonable performance requirements do exist, and the present system seems adequate to the task.

In the control system environment the efficiency of database derived transactions is clearly more important, and during the commissioning of the AGS to RHIC transfer line currently underway, it was decided that online database access would not be a central element in the control system. However, limited online database access to front end configuration data is now being tested and may be introduced in the next version of the engineering support software. If no inefficiencies are observed, the database will probably move into a more central role in the control system as RHIC approaches completion. On the other hand the AGS control system has been using Borland's InterBase[1] database management system successfully for many years. Front end system configuration data are stored and accessed, and some online transaction updating is performed.

The data design and access tools at RHIC are heterogeneous with a strong emphasis on public domain tools. For schema design we have been using *Erdraw* and the Object-Protocol Model, *OPM*, tools from the LBL data management group[4]. In addition, we have *PowerBuilder* for deploying data entry forms to PC's. For program applications Sybase supplies the Open Client C libraries, and extensive use of these routines has been made. Since the control system software for RHIC is being written in C++, these libraries have been organized into C++ class libraries[10]. If there is interest, these libraries can be made available publicly. Other publicly available share ware that is being used includes the *sybtcl*[11] toolset which allows the use of SQL within the Tool Command Language (TCL). This package also includes a generic X11 interface, *wisql*. This point and click interface is very convenient for users with little SQL experience. Finally, Sybase tools for the Linux OS have recently become available as Sybase has compiled their Open Client libraries under Linux and made them available at no charge[12].

Database access via the World Wide Web (i.e., through HTML interfaces) has also been implemented. The AGS to RHIC transfer line page at URL http://www.rhichome.bnl.gov/RHIC/ATR/ has many examples of reports generated directly from the database server using SQL scripts invoked from HTML. We do not currently support interactive updating or data entry from HTML interfaces. However, we are investigating the use of *SybPerl*[13] as a means of simplifying and expanding Web access to the database.

3 Office of Environmental Restoration Geographical Information System

BNL has been selected as a Superfund site because of potential tritium contamination of the ground water within the laboratory grounds and the land southward between the lab and the Great South Bay of Long Island. The Office of Environmental Restoration at BNL is responsible for conducting this effort. Wells are being dug on and off-site, and local homeowners' wells are monitored on a regular basis. The information from these studies is provided to the lab by Suffolk county as well as through contract workers involved in this project. The data is then loaded into an Oracle database system running

on a SUN sparc20 platform with about 20 GB of disk storage. The information recorded includes for example the well water measurements and the geographical coordinates of the sample site. The latter measurement provides the link to extensive information about the properties being studied so that the local government and citizenry may have many points of reference and views into the data.

A proprietary graphical information system called *Arc/Info*[5] is used to make geographical images of the areas onto which views of the sample data may be mapped. *Arc/Info* provides multi-color views of the sample measurements, and can respond to programmable database queries. The *Arc/Info*/Oracle system has been under development for about a year and a half, and will eventually be released to the BNL library as a publicly available database/geographical archive. Many sample sites are still being prepared for measurement so that the total data store is not yet complete.

4 Protein Data Bank

The Chemistry department at Brookhaven has maintained an important element of the Human Genome Project called the Protein Data Bank[15] for about 25 years, and during the last 4 years the system has been moved to a relational database system. The Protein Data Bank is an archive of experimentally determined three-dimensional structures of proteins, nucleic acids and other biological macromolecules. Chemistry is using Sybase on an eight processor Silicon Graphics server to manage this effort.

Since the data bank is to be shared with collaborators throughout the world, one of the principal reporting interfaces has been through the use of tools for the World Wide Web. Web/Genera[14] is one such tool. It is an active HTML interface to data objects for Sybase. Since Sybase (like all relational databases) is built around the use of tables, it cannot support "object" data structures directly. One first has to define object schemas using Genera's schema notation. Genera then interprets this notation in terms of Sybase stored procedures and formats an HMTL interface that invokes the objects/procedures. The output from the queries is also formatted in HTML.

Chemistry is also using the Object-Protocol Model (OPM) of Chen and Markowitz[4] to develop database designs in terms of object schemas that are then implemented in relational databases like Sybase. *OPM* was invented initially to respond to difficulties in modeling genome sequencing experimental procedures and their associated data. However, the data and methods entailed in performing generic scientific experiments can be described within *OPM*, and its object-oriented approach to data structure modeling is very intuitive. The Data Management Group at LBL has built a X11-Motif based user interface to develop schema designs and other tools that implement these schemas in SQL. There are also tools that extract the objects from the database for reporting purposes. These tools are publicly available by contacting Victor Markowitz[4].

5 Future Plans

5.1 RHIC & AGS

The AGS intends to change its RDBMS to Sybase from InterBase during the next year. The possibility of using Sybase's *Replication Server* is being considered. As the name suggests, *Replication Server* provides a "hot" backup system for critical DBMS's. RHIC is also discussing the possible use of an Object DBMS as a layer to record the activity of front end systems in a uniform way for reporting to the high level applications.

5.2 Protein Data Bank

Sybperl[13], an extension to *Perl5* using Sybase's Open Client libraries, is being used to upgrade the present WWW interfaces to the data bank.

5.3 NSLS

The National Synchrotron Light Source has begun to use Informix as part of its upgrade to the control system. Prototype data structures have been constructed and are being tested.

6 Conclusions

Relational databases are a crucial element in most of the information systems at Brookhaven. However, tools to access and update these systems are not as easy to build or as robust as most end users would wish. The popularity of the World Wide Web offers another opportunity to build cross-platform database interfaces. WWW interface tools that provide relatively simple HTML programming of database queries, data entry forms and reporting are needed. Using Perl (i.e., *OraPerl, SybPerl*, etc.) to produce HTML formatted reports directly from database queries is probably the quickest and cheapest way to integrate the Web with database access.

REFERENCES

[1] Borland International, Inc. 100 Borland Way, P.O. Box 660001, Scotts Valley, California, 95067-0001.

[2] Sybase, Inc., 6475 Christie Street, Emeryville, California 94608-9890.

[3] InSync Software Corporation, Ronkonkoma, New York.

[4] Chen, I. A. and Markowitz, V. A., *An Overview of the Object-Protocol Model (OPM) and OPM Data Management Tools*, Information Systems, Vol. 20, No. 5, pp. 393-418, Pergamon, 1995. See also http://gizmo.lbl.gov/DM.html

[5] Environmental Systems Research Institute, Inc. 380 New York Street, Redlands, California 92373-8100.

[6] Oracle Corporation, 500 Oracle Parkway, Redwood Shores, California 94065.

[7] S. Peggs et. al., *Relational Databases for RHIC Design and Control*, EPAC, 1994.

[8] C. G. Trahern et. al., *Design of the Commissioning Software of the AGS to RHIC Transfer Line*, PAC, Dallas, 1995.

[9] J. Wei et. al., *Field Quality Evaluation of the Superconducting Magnets of the Relativistic Heavy Ion Collider*, PAC, Dallas, 1995.

[10] Steve Tepikian, *A Reference Manual for the DB++ Class Library*, RHIC/AP/77.

[11] Tom Poindexter and U S WEST Enhanced Services, Inc.

[12] See URL http://www.sybase.com/Offerings/Samples/Linux/

[13] Michael Peppler, ITF Management SA, and See http://www.sybase.com/WWW/Sybperl/index.html

[14] S. Letovsky, M. Berlyn, *1995 Genera: A Specification-Drive Web/Database Gateway*, submitted to WWW '95. See also http://gdbdoc.gdb.org/letovsky/genera/genera.html

[15] See URL http://www.pdb.bnl.gov/

DISCLAIMER