ANL-HEP-CP-94-37

Conf-940492--10

The PASS Project: A Progress Report[1]

D. R. Quarrie[2], C. T. Day, S. Loken, J. F. Macfarlane
Information and Computing Science Division
Lawrence Berkeley Laboratory

D. Lifka, E. Lusk, D. Malon, E. May, L. E. Price
High Energy Physics, Mathematics and Computer Science,
and Decision and Information Science Divisions
Argonne National Laboratory

L. Cormell, A. Gauthier, P. Liebold, J. Hilgart,
D. Liu, J. Marstaller, U. Nixdorf, T. Song
Physics Research Division
Superconducting Supercollider Laboratory

R. Grossman, X. Qin, D. Valsamis, M. Wu, W. Xu
Laboratory for Advanced Computing and
Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago

A. Baden
Department of Physics
University of Maryland

ABSTRACT

The PASS project has as its goal the implementation of solutions to the foreseen data access problems of the next generation of scientific experiments. It is in the process of transitioning from an exploratory phase, where the focus has been on understanding the requirements and available technologies to an implementation phase, where detailed design work is commencing on a common framework for scientific applications.

INTRODUCTION

The Petabyte Access and Storage Solutions (PASS) project [1] has as its goal the implementation of solutions to the foreseen data access problems of the next generation of scientific experiments. These are characterized by a very large sample of complex event data (~10^{15} bytes), a dilute signal and a large (~1000) and geographically dispersed user community. Although the original focus of the PASS project was the experiments at the SSCL, its approach is broad enough to encompass many areas of scientific computing. Target customers now include experiments at RHIC, CEBAF, the B-Factory at SLAC, NASA and governmental projects.

The general approach has been to investigate the feasibility of using distributed database technology in conjunction with

1. Supported by the U.S. Department of Energy under Contract No. DE-AC03-76F00098.

2. Address: LBL MS50B-3238, 1 Cyclotron Road, Berkeley, CA 94720
Email: DRQuarrie@LBL.Gov

hierarchical mass stores to handle the conflicting requirements for the storage of large amounts of experimental data with the desire to provide rapid access to selected components of that data from a dispersed user community.

ORGANIZATION

PASS has been organized in two distinct phases, where the initial phase has been mainly an exploratory one, with the focus being on benchmarks and technology demonstrations to demonstrate proof of principle and understanding of the available hardware and software technology. The culmination of this initial phase has been the development of an Architectural Model that forms the basis for the second, implementation, phase. The second phase is focused on the detailed design and implementation of components identified by the Architectural Model. The goal is this phase is the creation of second-generation prototypes, eventually leading to systems capable of handling the data access demands for a wide variety of scientific disciplines.

The exploratory phase has involved two generations of testbeds, the first of which demonstrated that the database approach was feasible. The second generation testbeds have extended the size of the data sample, the complexity of the queries and have embarked on an investigation of access to distributed data and distributed queries.

MARK 0 TESTBED

The first generation testbed was described at an earlier CHEP Conference [2]. It demonstrated the feasibility of the database approach, and indicated that an object oriented approach using either an object oriented database manager or persistent object store (or a combination of the two) was the preferred approach. However, it was limited in the size of the data sample and complexity of the physics queries that could be performed.

The Mark 0 testbed indicated that either a full-scale object oriented database or an object persistence manager with lower overheads were suitable candidates for further investigation. PTool [3] is a persistent object manager developed at the University of Illinois at Chicago. A 32-bit version was used during the Mark 0 tests, whereas a 64-bit version with significantly enhanced capabilities has since been developed.

MARK 1 TESTBEDS

Several Mark 1 testbeds are underway. A testbed at the SSCL was designed to demonstrate the use of 19mm helical scan tape technology within a database environment and to increase the data sample to 10GB. The configuration is shown in Fig. 1. Two

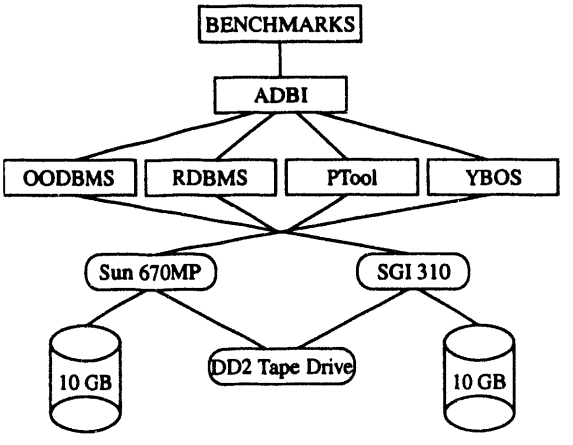

Figure 1. SSCL Configuration

different computer architectures were investigated as well as the same set of databases as for the Mark 0 tests, but with a significantly enlarged data sample. The demise of the SSCL prevented completion of these tests.

Another testbed has focused on the use of the 128-processor IBM SP-1 computer system at ANL and the development of parallel query processing techniques using the PTool persistent object manager. Several replicated query strategies have been investigated as well as techniques for the movement of data within a distributed database

environment. This project is the subject of a paper to be presented at this Conference [5].

Work at UIC has been targeted towards the further development of the PTool persistent object store into a fully distributed environment. This has allowed tests to be made with different caching, migration, and replication algorithms for interfacing low overhead, high performance persistent object managers to hierarchical storage systems. This work is described in Ref. [6].

Technical difficulties limited the complexity of the queries that were possible on all the above testbeds, so another testbed at LBL has attempted to integrate an existing physics analysis framework with a distributed OODBMS so that a larger sample of data may be examined and typical physics queries may be run. The CDF analysis framework has been modified to allow the OODBMS to become the source of event data, whilst allowing the user code to remain unmodified. This framework has further been enhanced to act as a testbed for the use of a distributed database based on the concepts of the Object Management Group [7] and the Common Object Request Broker Architecture (CORBA) which forms the basis for the Architectural Model described in the next section. This testbed is described in Ref. [8].

ARCHITECTURAL MODEL

Focal point of the exploratory phase of the project has been the development of an Architectural Model [9]. It is comprised of four major components:

(a) The operational and technical requirements. Operational requirements are broad capabilities that result from the environment within which the system must operate. Characteristics which drive the operational requirements are the high input bandwidth, the very dilute signal and the widely dispersed scientific community. Technical requirements a specific capabilities that the system must exhibit in order to match the operational requirements. These include the input bandwidth, uniformity and scalability constraints, flexibility and extensibility in the data organization, modes and patterns of access, concurrency and access controls, and query language.

(b) An abstract reference model. This describes a system that matches the requirements in terms of its components and the mechanisms by which they communicate, but does not discuss policy or management issues that would be necessary to match the model to an actual implementation. This reference model builds upon the concepts and terminology of several standards organizations, including the Object Management Group and the Object Database Management Group [10].

(c) An implementation model. This describes a conceptual implementation, matched to the requirements of a HEP collider experiment. It consists of a set of hierarchical data servers.

(d) A discussion of some design and policy issues. The reference model lacks many of the characteristics of a final implementation that accommodate technological constraints to optimize the available capacities. For example, the reference model states that data must be movable amongst a hierarchy of data stores in a manner that optimizes response times to the most frequent access patterns. How best to achieve this caching and migration, and whether to replicate or move the data, is a detailed policy and design issue that lies beyond the scope of the reference model. We have identified several such issues that are worthy of more discussion and have presented aspects of their impact, even though at this stage in the design process we cannot necessarily identify the correct design choice.

SUMMARY

This paper presents an overview of the PASS project, summarizing the results from the various testbeds and presenting a brief description of the Architectural Model. The Mark 0 tests showed the desirability of analyzing events using distributed database and distributed object computing technologies. The Mark 1 tests showed how this technology could be interfaced to hierarchical storage systems resulting in our current architectural model. We are now ready to scale this technology up to the production level.

We have embarked on a detailed design of an implementation of the Architectural Model and the software products developed thus far are being retrofitted to conform. In the short time frame, work is underway to populate object stores with D0 data at Fermilab and the longer term plan encompasses experiments at both SLAC and RHIC with the goal being the availability of production systems in the LHC time frame.

REFERENCES

[1] A. Baden and R. Grossman, *A Model for Computing at the SSC*, Superconducting Super Collider Laboratory Technical Report No. 288, 1990.

[2] C. T. Day et al., *Database Computing in HEP -- Progress Report*, Computing in High Energy Physics, 1992.

[3] R. Grossman and X. Qin, *PTool: A Software Tool for Working with Persistent Data*, Laboratory for Advanced Computing Technical Report Number 92-11, University of Illinois at Chicago, 1992.

[4] A. Gauthier et al., *PASS Project at the SSC: Test Plan for 10 Gb Database Comparison*, PASS Project Note 93-04, April 1993.

[5] D. Malon et al., *Parallel Query Processing for Event Store Data*, to be presented at CHEP '94.

[6] E. N. May et al., *A Demonstration of a Multi-level Object Store and its application to the Analysis of High Energy Physics Data*, to be presented at CHEP '94.

[7] Object Management Group, *The Common Object Request Broker: Architecture and Specification, Revision 1.1*, OMG TC Document 91.12.1, 1991.

[8] D. R. Quarrie and C. T. Day, *Using a Distributed OODBMS as a Source of Events for CDF Physics Analysis*, to be presented at CHEP '94.

[9] D. Lifka et al. (the PASS Collaboration), *The PASS Project Architectural Model*, to be published.

[10] R. G. G. Cattell et al., *The Object Database Standard: ODMG-93*, Morgan Kaufmann, 1994.