Conf-940492--9
ANL-HEP-CP-94-38

The PASS Project Architectural Model*

C. T. Day†, S. Loken, J. F. Macfarlane and D. R. Quarrie
Software Technologies and Applications Group
Information and Computing Sciences Division
Lawrence Berkeley Laboratory

D. Lifka, E. Lusk, D. Malon, E. May, and 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. The architectural model results from an evaluation of the operational and technical requirements and is described in terms of an abstract reference model, an implementation model and a discussion of some design aspects. The abstract reference model 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 design issues that would be necessary to match the model to an actual implementation. Some of these issues are discussed, but more detailed design and simulation work will be necessary before choices can be made.

---

*Supported by the U. S. Department of Energy under Contract No. DE-AC03-76FS00098.

†Address: Lawrence Berkeley Laboratory, MS 50B-3238, 1 Cyclotron Road, Berkeley, CA 94720.
Email: CTDay@LBL.Gov.

INTRODUCTION

The Petabyte Access and Storage Solutions (PASS) project [1], [2] has as its goal the implementation of solutions to the foreseen data access problems of the next generations of scientific experiments. These are characterized by a very large sample of complex event data, a dilute signal and a large and geographically dispersed user community.

We describe an Architectural Model that is comprised of four major components: the operational and technical requirements, an abstract reference model, an implementation model, and some design issues.

The architectural model is the basis of the transformation of the PASS project from its initial exploratory phase to an implementation phase where components of the model are being designed and implemented.

REQUIREMENTS

The requirements are presented as both operational and technical. Operational requirements are broad capabilities that result from the environment within which the system must operate. We have used the model of an HEP collider detector as the basis for these, although most would be similar in other areas of scientific research such as nuclear physics, satellite telemetry and high resolution astronomical surveys. Of prime importance for the operational requirements are that the input bandwidth is high, that the signal is very dilute and that the scientific community is widely dispersed.

Technical requirements are 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, concurrency and access controls, and a query language.

THE REFERENCE MODEL

We have based the reference model on the concepts and terminology from commercial organizations [3, 4, 5]. By this means we hope to achieve the maximum leverage to be able to incorporate components from a large variety of sources. Within different application domains different sets of components might need to be present, but they would share significant commonality and would be easily replaceable within the same common framework.

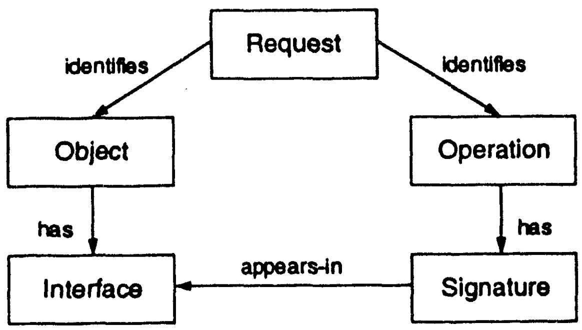

Figure 1. Object Semantics

The reference model is presented from a variety of different viewpoints that reflect the overall complexity of the system and highlight different aspects of it. The information viewpoint concentrates on information modeling, structure and flow. It is primarily based on the Object Management Group Object Model [3]. Figure 1 is a representative diagram from this level.

The computational viewpoint focuses on interactions between the entities identified within the information viewpoint and their exchange of data and control. It is expressed in terms of Object Services as in the OMG's Service Architecture [5].

The engineering viewpoint deals with the communication and support mechanisms; objects outside of the nominal application domain make their appearance here. Much of this viewpoint follows the OMG's CORBA [4]. However, Figure 2 shows an additional abstract mechanism we have identified to make distributed services transparent to client objects [6].

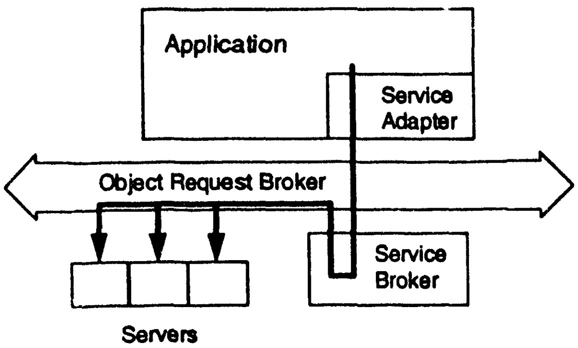

Figure 2: A Brokered Service

The technology viewpoint concentrates on the impact of available technology on components of the complete model. An example of this viewpoint is presented which connects our model with the IEEE Mass Storage Model.

AN IMPLEMENTATION MODEL

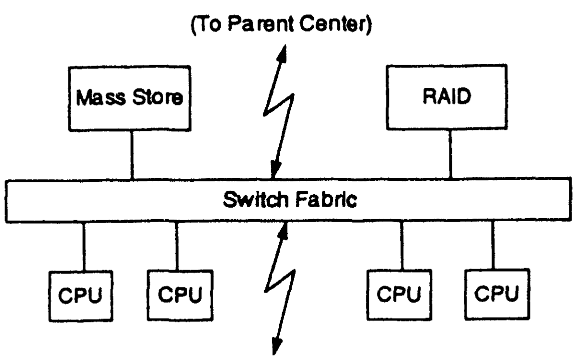

Figure 3. Generic Data Center

The implementation model focuses mainly on hardware aspects of a solution, including capabilities such as network bandwidth and mass store capacities and bandwidth. We present an implementation model made of a geographically dispersed hierarchy of Computing Center Nodes. A typical node is shown in Figure 3.

DESIGN ISSUES

A reference model lacks many of the characteristics of a final implementation that accommodates technological constraints to optimize the available capabilities. For example, the reference model states that data must be moveable 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 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 presented aspects of their impact.

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] Object Management Group, *Object Management Architecture Guide, Revision 1.0*, OMG TC Document 90.9.1, 1990.

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

[5] Object Management Group, *Object Services Architecture, Revision 6.0*, OMG TC Document 92.8.4, 1992.

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