MASS STORAGE & DATA MANAGEMENT

Jamie Shiers

Computer Division, CERN, CH-1211 Geneva, Switzerland

The Mass Storage and Data Management parallel session covered submissions on data storage technologies, data management systems, databases and their use in HEP applications. Rather than summarise each paper, this session summary attempts to identify the trends in HEP data management and mass storage.

Hep Requirements

Hep experiments have a number of requirements that are somewhat atypical in the area of mass storage and data management. These are the sheer volume of data involved - already at the level of tens of TB for existing experiments, rising through hundreds of TB per year for a number of experiments starting around 1999, to one or more PB per experiment per year at the LHC. The total data sample integrated over the projected LHC running period and all experiments could be as much as 100PB - significantly more than is typically involved in other domains. Coupled with these enormous data volumes are large data rates - 100MB/second to over 1GB/second for LHC experiments.

Fortunately, there are a number of simplifying characteristics of Hep data. These are that the data is very largely read-only, that the number of simultaneous users is low - very low compared to figures seen in many database environments. Similarly, the "transaction rate" is also very low.

Mass Storage Systems

The IEEE Computer Society has been developing a reference model for Mass Storage Systems over many years. A number of commercial systems, based upon various versions of this reference model, have been used in the HEP community with varying degrees of success over the past years. DESY, KEK and Thomas Jefferson laboratory (CEBAF) have all successfully deployed mass storage solutions based upon the Lachman (later Legent then Computer Associates) Open Storage Manager product. This system has been used in production at DESY for a number of years and has now been interfaced to DFS at KEK.

More recently, a collaboration of laboratories and commercial companies in the US has developed a system known as HPSS (High Performance Storage System). This system, which owes much to the IEEE reference model and to previous work performed at large US laboratories, including CFS/Datatree from Los Alamos, Unitree from Lawrence Livermore National Laboratory and others, is the only known system designed to scale to the PB region. A number of HEP sites already deploy, or have plans to deploy, HPSS systems. These include Caltech, CERN, FNAL and SLAC. As such, it is fair to say that HPSS is emerging as the mass storage system of choice in the HEP community, although there are still a number of concerns about its as yet unproved reliability, and the fact that mass storage systems are unlikely to become, at least in the short term, true commodity products. This poses a question mark over their long term survival. Although it seems almost certain that some solution will exist, it may not be commercially supported nor easily affordable if there is insufficient market to sustain it. For comparison, the most successful mass storage system ever deployed, namely CFS, was only installed in some 20 sites, none of which paid to acquire it, although many did pay for maintenance and support.

Data Management

Data management has long been a key area in HEP computing, requiring significant manpower resources to provide largely home-grown solutions. Data management in HEP has finally outgrown the traditional file and tape based solutions previously adopted, with most experiments deploying event directories, logical datasets and so forth. Until recently, however, these systems have been largely incompatible and inflexible. Current developments strongly suggest that object level access to event data is realisable with existing solutions, commercial or otherwise. Such solutions promise to revolutionise HEP data management, solving the problem of "access to any byte in a PB", first discussed at CHEP '92.

As well as offering a great deal of flexibility, several independent measurements of both commercial ODBMS-based solutions and home-grown systems have shown that object-based access offers performance that is at least equal to, and often, particularly in the case of random access with sufficiently low selectivity, much better than that offered by traditional systems.

Summary

There is a general trend in the HEP community to move towards commercial mass storage systems based upon the IEEE Computer Society Reference Model for Mass Storage Systems. In particular, HPSS appears to be the preferred system for future HEP experiments and is already deployed at a number of sites, with installations planned for a number of others in the short to medium term.

File-based data management solutions are becoming a thing of the past, being replaced by object-based solutions. Memory management is now handled by the programming language, although object persistency still requires an additional system. Here, ODMG-compliant solutions appear to be gaining ground and commercial solutions have been used in production on a number of experiments.

Plans for interfaces between Objectivity/DB, the ODBMS product currently studied by RD45, and HPSS should result in a highly scalable, distributed solution capable of meeting the requirements of experiments starting around 1999, as well as addressing the problems of the LHC collaborations. The first prototypes of such solutions are expected towards the end of 1997, with production versions approximately one year later.