B DATA ACCESS AND STORAGE

B.1 DATA STORAGE TECHNOLOGY

ABS_106

A Pilot Mass-Storage system for KEK Belle Experiment

H. FUJII (KEK)
R. Itoh, A. Manabe, A. Miyamoto, Y. Morita, T. Nozaki, T. Sasaki, Y. Watase and T. Yamasaki (KEK and Sony)

A pilot mass-storage system for KEK Belle (B-meson Physics) experiment has been developed. This experiment requires a high speed and large data recording system. The required recorinding speed is about 3 MB/sec in average and 15 MB/sec at maximum. The requried volume is more than 30 TB/year. We have developed a pilot system to study the high-speed and large volume data recording system which satisfies the above requirements. The system consistes of (1) SONY-DIR 1000M data recorder with SCSI-2 Fast/Wide interface; the recording capability of the device is 16MB/sec at maximum. (2) SONY-DMS 24 tape robot (tape library) of which volume capacity is about 2 TB (3) high-speed TCP/IP network of HIPPI and (4) three workstations running under UNIX. For the system software, the CERN-SHIFT system has been installed for the study. Because that the tape device and the robot (tape library) system are completely different from currently running SHIFT systems, we have rewritten the tape server part. The tape device and the robot(library) control path are directly connected to UNIX workstations. To achieve the required recording speed, we also developed an application interfaces for this tape server. We have made the user interface without using tape staging mechanism. This user interface reduces the overhead of the recording system which is critical for the Belle expriment. This tape server system has developed based on TCP/IP, so that the system is easy to expand and free from network media.

ABS_73

Object Based Data Access at The DO Experiment

S. FUESS (FERMILAB)
S. Ahn, M. Galli, L. Lueking, L. Paterno (Fermilab), H. Li (Stony Brook)

The DO Experiment at Fermilab is finishing approximately three years of data taking at the Tevatron. Over 30 Terabytes of reconstructed data have been accumulated. In the next collider run, scheduled to begin circa 1999, DO expects to collect approximately 160 Terabytes of data per year. The access to these vast quantities of data is one of the greatest problems of experimental data analysis. DO is currently participating in the Fermilab Computing Division "Computing for Analysis Project" (CAP) to investigate a solution using object based data storage and access. The CAP solution is to divide events into objects and independently store these objects on disk or tape. Disk storage is on an IBM SP-2 system, allowing high speed parallel access. A IBM 3494 robotic tape library provides the bulk of the storage.

DO has taken its current ZEBRA data model and produced a corresponding object model. Data has been imported from ZEBRA into object stores, which are subsequently available for data mining activities. Control of the scanning is achieved via a query language accessible on a World Wide Web page, and is further customizable with an explicit user-supplied FORTRAN, C, or C++ function. Output of a scan may be kept in object format or exported back into ZEBRA format. We will present details of the model, progress to date, and future plans.

ABS_70

Study for possible storage solutions for KLOE

M.L. FERRER (FRASCATI)
F.Donno, E.Pace (Frascati)

The requirements of KLOE represent a challenge in this field, since we have to deal with data rates (about 10**11 events/year) considerably higher than those of typical collider experiments such as LEP or TEV-1, albeit a factor ten smaller then those foreseen at the LHC. It is estimeted that the average event size is 5 kBytes, corresponding to a total bandwidth requirement of 50 MBytes/sec. The amount of data collected each year of running is then of the order of 500 TBytes. The total storage requirements are even greater due to the MonteCarlo data that will be required and could only be managed with special robotics and a good computer organization. In this paper we discuss the characteristics of actual tape drive systems and robotic libraries that could be of interest for KLOE (DLT, IBM 3590, STK Redwood, AMPEX DST 310) and results of test are shown for DLT's (under HP UX and Digital UNIX). The software that will be used for

reading/writing data is also described. Home made software will be employed to drive the robotic parts.

ABS_56

Data Access and Management using the DMI Filesystem Interface

M. GASTHUBER (DESY)

Several months ago DESY started using a new scheme to access the physics data from UNIX based batch machines using the hierarchical Mass-Storage system called OSM (Open Storage Manager). This system provides a network based, file-oriented put/get repository together with a simple set of user commands for access and management. The applications usually requires random access, where the Mass-Storage system only supports a sequential byte stream. This fact leads to the use of dedicated staging systems. These staging systems seem to be sufficient in the case where enough disk (stage) space is available and/or the staging rate is moderate. To allow a better disk-cache efficiency one need to have extended knowledge about the access statistics which recently is only available through the UNIX filesystem. To incorporate site specific access policies this tight connection to the filesystem is also required. The connection between a migration capable filesystems and the hierarchical Mass-Storage system is the first goal. The second goal is the ability to manage large amount of diskspace, distributed between various machines and building something like a 'data-domain' while still having a high performance distributed data access. The Data Management Interface (DMI) recently becomes available and offers all requested functionalities for efficient filesystem controlling and management. It allows standard, kernel independent filesystem control and IO operations and will be available on various UNIX flavors soon.

This talk will show the design and first practice of such a system using Silicon Graphics DMI implementation and the OSM as the hierarchical Mass-Storage system managing large tape libraries.

ABS_25

Data Access and Data Storage in H1

R. GERHARDS (DESY)
U. Kruener-Marquis (DESY)

The electron proton collider HERA at the DESY laboratory in Hamburg and the H1 experiment are now in successful operation for more then three years. Electrons and protons collide at a bunch crossing rate of 10.41 Mhz, placing demanding requirements on the data acquisition system but also on the data storage and data access.

The H1 experiment is logging data at an average rate of 500 KB/sec which results in a yearly raw data volume of several Terabytes (e.g. 2.5 TB in 1994). The data are reconstructed with a delay of only a few hours, yielding another Terabytes of reconstructed data after physics oriented event classification. Both the data logging and the quasi-online event reconstruction are performed on an multiprocessor SGI Challenge computer which is in particular providing a very powerful I/O performance. An Ampex DST 800 mass storage system with 3 DST 600 recorders has been connected to this Challenge to store both the raw and reconstructed data. This allows much faster and more convenient access to the data at rate of up 10 MB/sec.

Physics analysis is performed on another SGI Challenge computer, equipped with about 500 GB of disk and, since a couple of months, direct access to a StorageTek ACS 4400 silo, using the Open Storage Manager (OSM) software. The disk space is mainly devoted to store the reconstructed data in very compressed format (typically 5 to 10 KB per event). This allows for very efficient and fast physics analysis which needs random access to the data. Monte Carlo data, on the other hand, are kept in the ACS silo and staged to disk on demand.

A detailed report will be given on the experience with this system in a demanding comuting environment at HERA.

B.2 DATA ACCESS SOFTWARE

ABS_101

The RITA network. How the high energy tools can be used in order to transmit clinical hadrontherapic data

S. SQUARCIA (GENOVA)
M. Ferrari, P. Risso (Genova)