Belle Computing Model

Nobu Katayama [a], Ryosuke Itoh [a], Atsushi Manabe [b] and Takashi Sasaki [b]

[a] *Physics Division*
High Energy Accelerator Research Organization (KEK)
Oho 1–1, Tsukuba-shi, 305, Japan

[b] *Computing Research Center*
High Energy Accelerator Research Organization (KEK)
Oho 1–1, Tsukuba-shi, 305, Japan

The Belle collaboration is now building a detector at the KEKB B factory at KEK in Japan. Its goal is to study the origin of CP violation. The experiment will be commissioned in 1999. In this paper the computing model of the Belle experiment is described.

Key words: CP violation; computing model;

1 Introduction

The goal of the Belle experiment[1] is to make a definitive test of the Standard Model's predictions for CP violations in the decays of B mesons. Determination of the unitarity triangle angles and sides requires studies with data sample containing $10^{8-9} B\bar{B}$ mesons. This, in tern, requires the use of very high current electron and positron beams and a detector system capable of handling high event rates and large data samples. The computing and software are of importance to the experiment as very complex data analyses using a large ammount of data is required for physics discoveries. The computing model represents the architecture of a system of the hardware and software. In the next section the requirements are discussed. Then the acquisition of the Belle computing system is explained. Finally the future plan will be discussed.

2 Computing and software requirements

The philosophy for the architecture for the Belle computing and software was as follows.

– The data path should be straight forward and simple.
– The mass storage system should be large enough to store the data taken by the Belle detector. The access to the data in the mass storage system should be fast enough for physics data analysis.
– The maintenance of the system should be done by a minimum number of people.
– The system should be easy to use for novice physicists.

In short the collaboration adopted the traditional HEP data/programming model.

With the above in mind we start with the requirements for the data acquisition process, proceed with data storage, processing and analysis.

The maximum data rate from the Belle detector is 15 MB per second as the maximum trigger rate is 500 Hz and the maximum event size 30 KB. The estimated raw data stored in the mass storage system is as large as 17 TB per year. At the design luminosity of $10^{34} cm^{-2} s^{-1}$ the event rate for hadronic events is 40 events per second. The raw data will be stored in a tape library of about 25 TB until the event reconstruction processing is done and the tapes may be moved out of the library into the storage area.

The traditional style data summary tapes (DST) will be produced from tape to tape running reconstruction algorithms on compute servers. The DST will be stored in a large tape library so that the DST can be accessed at anytime without having human operators mount the tape. The users may access DST if necessary but such situation is considered rare. The mini DST for final physics analysis will be produced for the users.

As the Belle detector is a general purpose detector like many other experiments at lepton colliders the mini DST for hadronic and QED events must be kept online. For this purpose disk space of 1.5 TB per year is required. The physics data analysis will be done using the mini DST on the disk for hadronic and QED events.

3 Belle computer system acquisition

The KEK computing research center (CRC) supports several computing systems for various divisions of KEK; Proton synchrotron, photon factory, TRISTAN, theory etc. The computer system for the experiments at TRISTAN was to be replaced with a new system dedicated for the users of the new KEKB B factory. As the Belle collaboration is the primary user of the system, the collaboration worked together with CRC for the acquisition.

As the acquisition is done by open international bidding, KEK writes a specification for the system and companies respond with proposals. The "best" proposal for "yen" is chosen.

At KEK the computer system that is as large as the Belle computer system must be leased from a single company for four years. The system is too large and complex for small companies to come up with heterogeneous (PC) solutions. The entire system must be leased from the company and be maintained to the specification 24 hours a day. Only large vendors such as DEC and IBM can respond to the biding.

The specification was written up by the members of the Belle collaboration and the CRC researchers. It took several months to come up with the final specification. In the specification it is required that

– the compute servers have computing power of more than 30,000 SPECInt92 in total.
– the tape library for the raw data have capacity at least 20 TB. The raw data must be written at 15 MB per second to catch up with the maximum DAQ data rate.
– the tape library for DST have capacity at least 70 TB and extendible to 120TB. The aggregate transfer speed between the compute servers and tape drives must be at least 81 MB per second.
– the tape library for HSM and the user data analysis have capacity at least 50 TB and extendible to 120TB. The aggregate transfer speed between the compute servers and tape drives must be at least 108 MB per second.
– the disk system have capacity larger than 2 TB. The aggregate transfer speed between the compute servers and disk system must be at least 200 MB per second.
– work group servers for interactive use have computing power of more than 9000 SPECInt92.
– . . .

Fujitsu Limited won the bid with the system[2] shown in Figure 1. The system consists of

– high speed data transfer system[3],
– a 27 TB tape library (SONY PeraSite DMS–8800) for the raw data using eleven SONY DIR–1000(M) tape drives,
– six units of Ultra Enterprise Server with 28 CPUs and one unit with 11 CPUs. The CPU is the 167MHz Ultra Sparc chip that has 256 SPECInt 92 of computing power,
– a 75 TB tape library (SONY PetaSite DMS–8400) for DST and production Monte Carlo data with 15 SONY DTF drives. (12MB/s),
– two 32 TB tape libraries (SONY DMS–8800) with 16 DTF drives. These

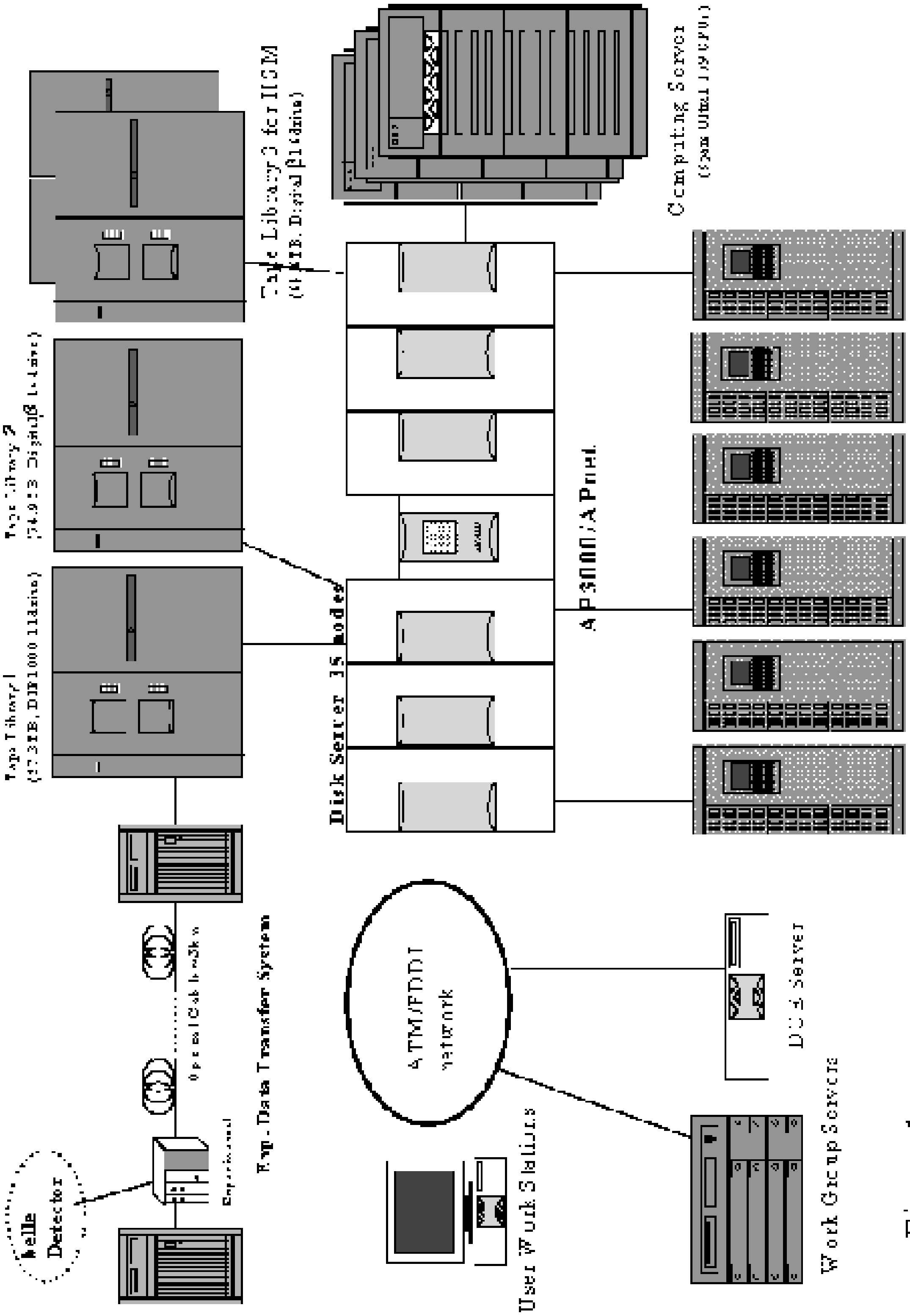

Fig. 1. The Belle computer system

will be used for backup and HSM as well.
– three work group servers with 28, 12 and 12 CPUs for interactive data analysis,
– six units of GEN5/XL RAID system totaling 4.2TB of online disk space,

The above elements are integrated as a Fujitsu AP3000 parallel server system. The compute and disk servers are connected via "AP-Net," a high speed network of transfer rate of 200 MB/s. The system is connected to desktop user workstations via ATM and FDDI LAN and X terminals via Ethernet.

4 Software

We have defined the following software components as baseline software on the system.

– Load Sharing Facility (LSF) batch system from Platform Computing corporation.
– DCE/DFS. On Solaris, Transarc corporation provides the software.
– Sony PetaServe HSM software (derived from OSM)[4]
– Home grown parallel data processing software within an SMP server (BASF)[5].
– Home grown I/O package (Panther).

Using these software we believe that the Belle computing model is fully supported. The LSF batch system will make effective use of seven compute servers. BASF enables us to use the SMP compute servers (with up to 28 CPUs) most effectively and the fully optimized I/O package (Panther) utilizes the fast IO channels to disk and tape drives from the compute and work group servers. We are hoping that DCE/DFS assures the secure and reliable data access from the collaborating institutes outside KEK.

5 Current Status

The entire system was installed by January 1997 at CRC. The software engineers from Fujitsu, researchers at CRC and members of the Belle software group are busy testing the hardware and baseline software. At the same time, the system has become available to the members of the Belle collaboration and other users of KEKB B factory. The migration from the old systems to the new computer system has been bumpy for physicists and will take several more months.

6 Future plans

The current Belle computing model is based on the traditional model of HEP computing. The choice was made earlier because of the time lag for the acquisition process of the computing system. The primary goal of the current model was a smooth and quick start of the experiment. We consider that this goal will be achieved.

The leasing contract forces us to go through the open bidding process every four year. Although it involves a lot of hard work it gives us a chance to shift from the traditional computing model to the new model. In particular the current serial access to incrementally refined data sets (raw, DST, mini DST and N-tuples) may not be the best. The common object store may be accessed using standard tools such as ODMG compliant object database.

The Belle software group will support the current computing model on the installed hardware for a smooth commissioning of the detector in 1999. However considering the experiment's long lifetime, the transition must come. The transition to the new programming languages, data access methods and tools must be planed carefully. We have switched our preferred programming language from FORTRAN to C++ last year. The home grown I/O system has incorporated the C++ interface based on STL. The home grown parallel processing system has been rewritten in C++ and has API for users of C++. A collaboration with the CERN RD45[6] and GEANT 4[7] projects are planed. Using the current RAID data disk system and SONY PetaServe migration software a prototype analysis system using Objectivity/DB database system is being designed. The collaboration may not achieve the true object orientation by the time of detector commissioning. However it is moving to the new paradigm cautiously but steadily.

References

[1] BELLE Technical Design Report, KEK Report 95–1, April 1995.

[2] R. Itoh *et al.* these proceedings, submitted paper f217.

[3] H. Fujii *et al.* these proceedings, submitted paper b313.

[4] A. Manabe *et al.* these proceedings, submitted paper c237.

[5] R. Itoh *et al.* these proceedings, submitted paper a244.

[6] A persistent Object Manager for HEP (Status Report)". CERN/LHCC,96-15, 1996.

[7] S. Giani *et al.* these proceedings, submitted paper.