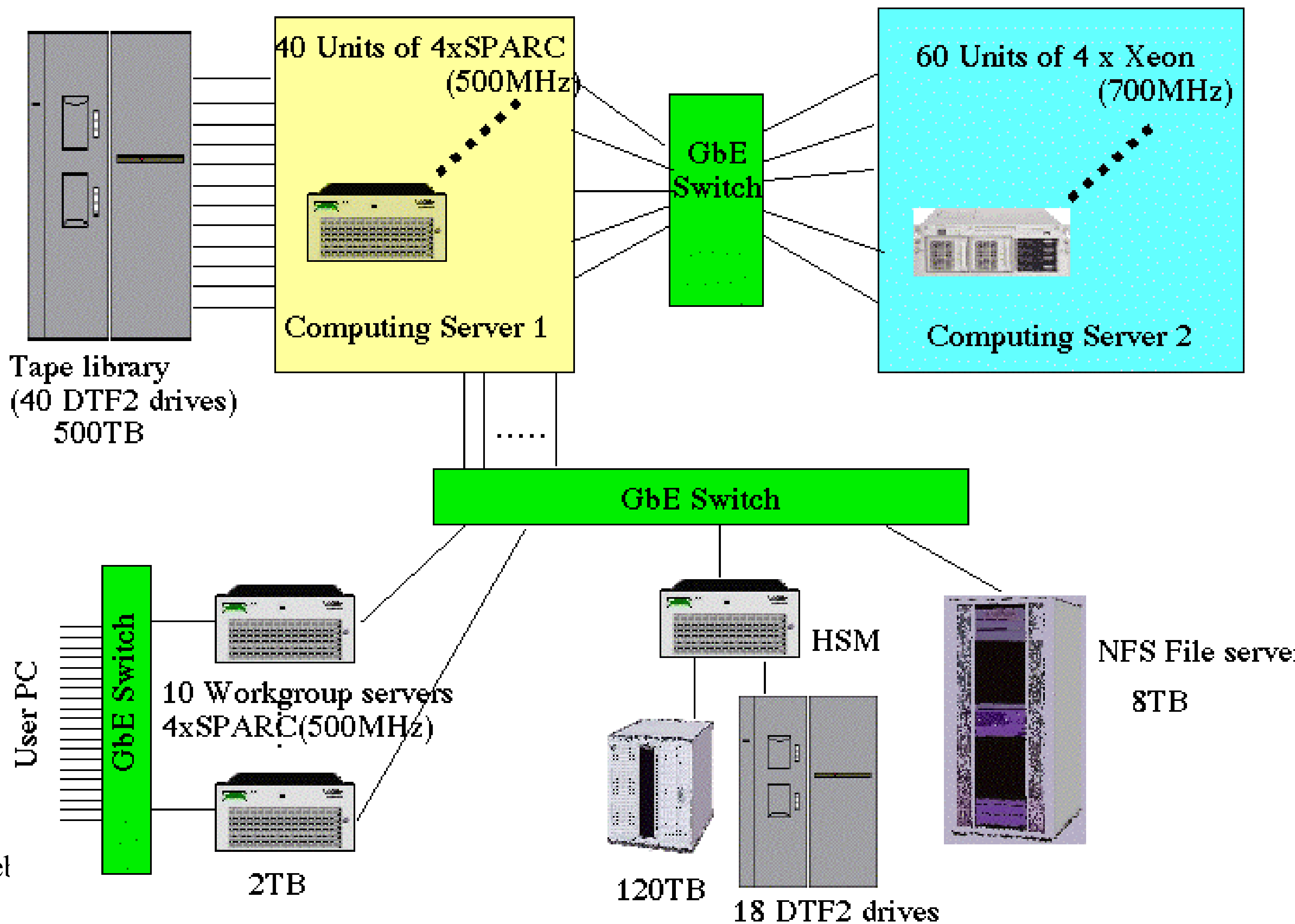
# Computing @ Belle

Belle-BaBar Detector Workshop

Nobu Katayama

KEK

# KEKB computer system




Fel

# Sparc CPUs

- Belle's reference platform
  - Solaris 2.7
- 9 workgroup servers (500Hz, 4CPU)
- 38 compute servers (500Hz, 4CPU)
  - LSF batch system
  - 40 tape drives (2 each on 20 servers)
- Fast access to disk servers
- 20 user workstations with DAT, DLT, AITs

# Intel CPUs

- Compute servers (@KEK, Linux RH 6.2/7.2)
  - 4 CPU (Pentium Xeon 500-700MHz) servers~96 units
  - 2 CPU (Pentium III 0.8~1.26GHz) servers~167 units
- User terminals (@KEK to log onto the group servers)
  - 106 PCs (~50Win2000+X window sw, ~60 Linux)
- User analysis PCs(@KEK, unmanaged)
- Compute/file servers at universities
  - A few to a few hundreds @ each institution
  - Used in generic MC production as well as physics analyses at each institution
  - Tau analysis center @ Nagoya U. for example

# Belle jargon, data sizes

- DST:120KB/hadronic event
- MDST:10(21)KB/hadronic(BBbar MC) event
  - zlib compressed, four vector+physics information only
- production/reprocess
  - Rerun all reconstruction code
  - reprocess:process ALL events using a new version of software
- generic MC
  - QQ (jetset c,u,d,s pairs/QQ generic b$\rightarrow$c decay) MC events for background study

# Data storage requirements

- Raw data 1GB/pb$^{-1}$ (160TB for 160fb$^{-1}$)
- DST:1.5GB/pb$^{-1}$/copy (240TB for 160fb$^{-1}$)
- Skims for calibration:1.3GB/pb$^{-1}$
- MDST:45GB/fb$^{-1}$ (7TB for 160fb$^{-1}$ )
- Other physics skims:~30GB/fb$^{-1}$
- Generic MC:MDST: ~10TB/year

# Disk servers@KEK

- 8TB NFS file servers
- 120TB HSM (4.5TB staging disk)
  - DST skims
  - User data files
- 500TB tape library (direct access)
  - 40 tape drives on 20 sparc servers
  - DTF2:200GB/tape, 24MB/s IO speed
  - Raw, DST files
  - generic MC files are stored and read by users(batch jobs)
- ~12TB local data disks on PCs
  - Not used efficiently at this point

# Software

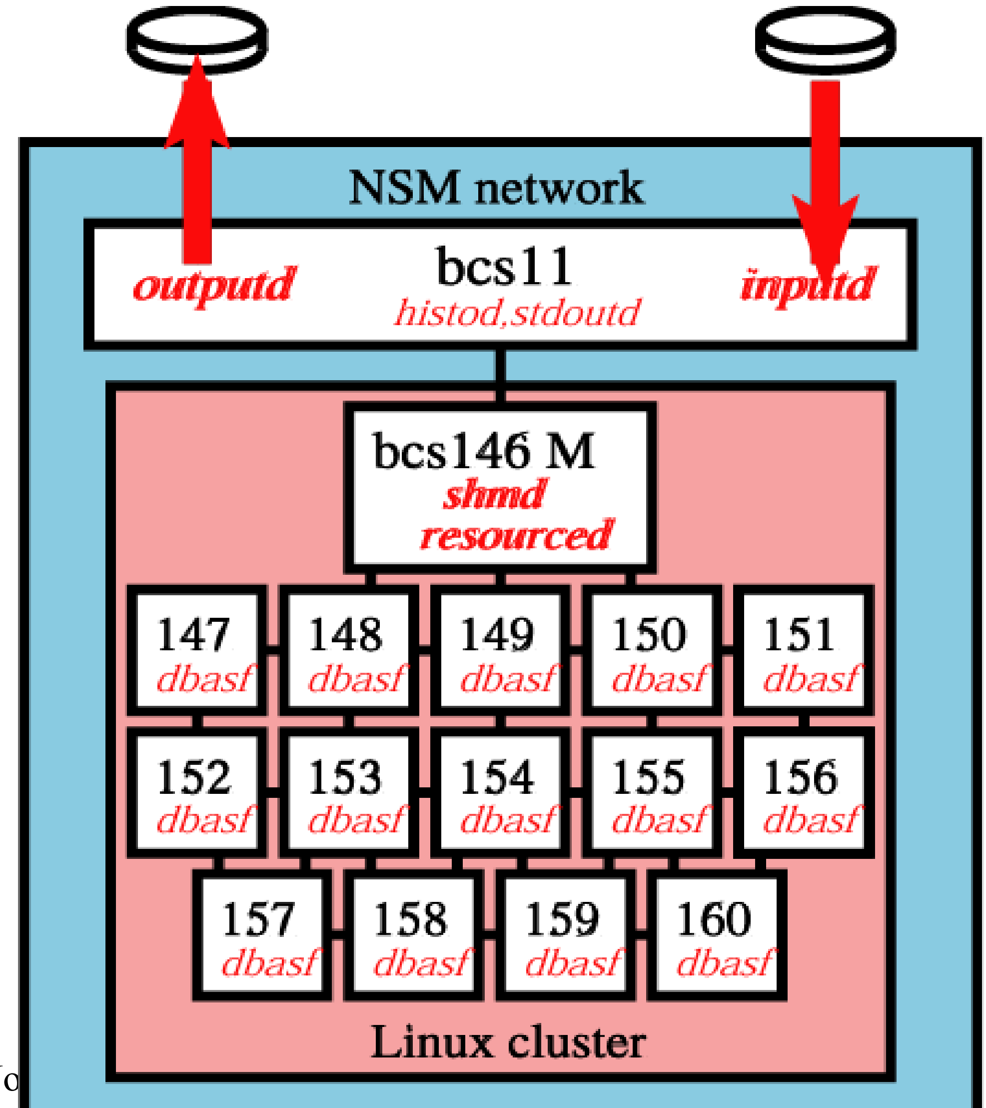
- C++
  - gcc 2.95.3 (compiles with gcc3, SunCC)
- No commercial software
  - QQ, (EvtGen), GEANT3, CERNLIB, CLHEP, postgres
- Legacy FORTRAN code
  - GSIM/GEANT3/ and old calibration/reconstruction code)
- I/O:home-grown serial IO package+zlib
  - The only data format for all stages (from DAQ to final user analysis skim files)
- Framework:Basf

# Framework (BASF)

- Event parallelism on SMP (1995~)
  - Using fork (for legacy Fortran common blocks)
- Event parallelism on multi-compute servers (dbasf, 2001~)
- Users' code/reconstruction code are dynamically loaded
- The only framework for all processing stages (from DAQ to final analysis)

# DST production cluster

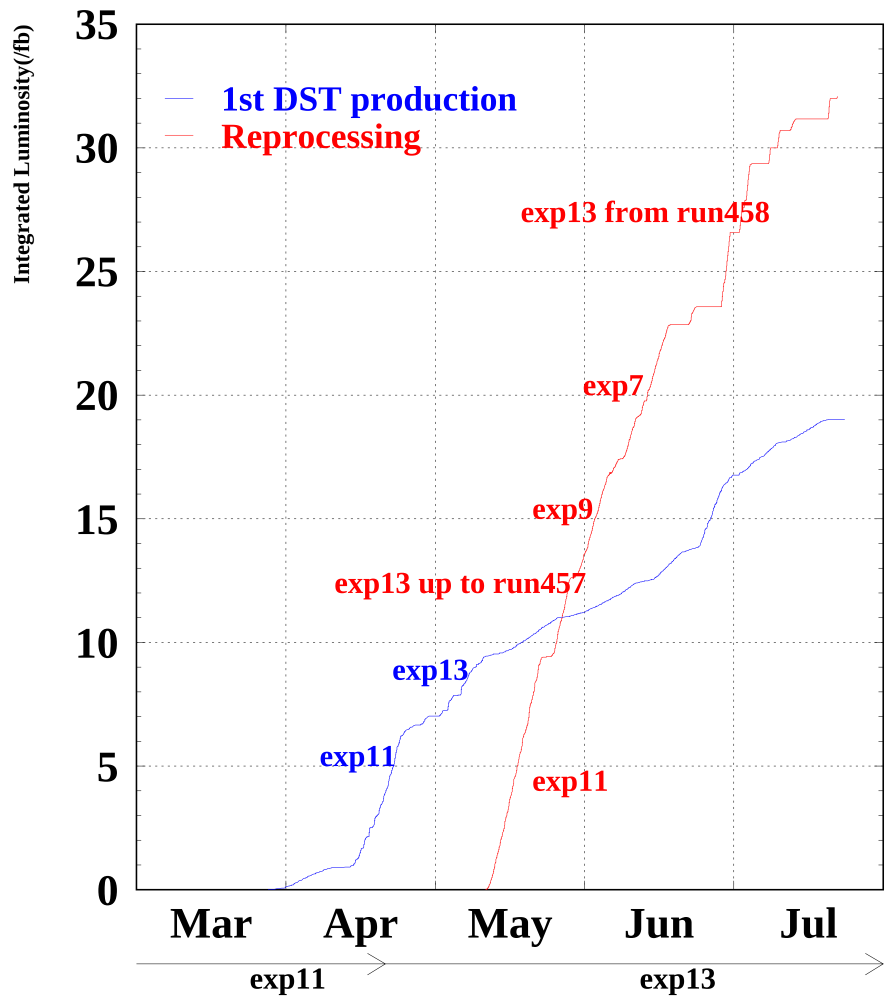
- I/O server is sparc
  - Input rate:2.5MB/s
- 15 compute servers/cluster
  - 4 700MHz Pen III Xeon
- 200 $pb^{-1}$/day
- Several such clusters may be used to process DST
- Using perl and postgres to manage production
- Overhead at the startup time
  - Wait for comunication
  - Database access
  - Need optimization
- Single output stream



# Belle Software Library

- CVS (no remote check in/out)
  - Check-ins are done by authorized persons
- A few releases (two major releases last year)
  - Usually it takes a few weeks to settle down after a release as we have no strict verification/confirmation procedure so far. It has been left to the developers to check the “new” version” of the code. We are now trying to establish a procedure to compare against old versions
  - All data are reprocessed/All generic MC are regenerated with a new major release of the software (at most once per year, though)

# DST production



# DST production

- 300GHz Pentium III~1fb$^{-1}$/day
- Need ~40 4CPU servers to keep up with data taking at this moment
- Reprocessing strategy
  – Goal:3 months to reprocess all data using all KEK compute servers
    - Often we have to wait for constants
    - Often we have to restart due to bad constants

# Skim

- Physics skim (MDST level)
  - Hadron, J/Ψ, Low multiplicity, $\tau$, $\eta_c$
- Calibration skim (DST level)
  - Hadron (A:loose cut), (C:very tight cut)
  - QED:Radiative Bhabha/ Bhabha/ Mupair, Radiative Mupair
  - Tau, Cosmic, Low multiplicity, Random

# Data quality monitor

- DQM (online data quality monitor)
  - run by run histograms for sub detectors
  - viewed by shifters and detector experts
- QAM (offline quality assurance monitor)
  - data quality monitor using DST outputs
  - From # of hits to reconstructed D* mass
  - WEB based
  - Viewed by detector experts and monitoring group
  - histograms, run dependence

# MC production

- 400GHz Pentium III~1$fb^{-1}$/day
- 240GB/$fb^{-1}$ data in the compressed format
- No intermediate (GEANT3 hits/raw) hits are kept.
  - When a new release of the library comes, we try to produce new generic MC sample
- For every real data taking run, we try to generate 3 times as many events as in the real run, taking
  - Run dependence
  - Detector background are taken from random trigger events of the run being simulated

  into account

# Postgres database system

- The only database system
  - other than simple UNIX files and directories
  - Moving from version 6 to 7
  - A few years ago, we were afraid that nobody use postgres but it seems postgres is the only database on Linux and is well maintained
- One master, one copy at KEK, many copies at institutions/on personal PCs
  - ~20 thousand records
  - IP profile is the largest/most popular
- The master hangs from time to time and makes a big headaches

# Reconstruction software

- 30~40 people have contributed in the last few years
- For most reconstruction software, we only have one package, except for muon identification software. Very little competition
  - Good and bad
- Identify weak points and ask someone improve it
  - Mostly organized within the sub detector groups
  - Physics motivated, though
- Systematic effort to improve tracking software but very slow progress

# Analysis software

- Several people have contributed
  - Kinematical and vertex fitter
  - Flavor tagging
  - Vertexing
  - Particle ID (Likelihood)
  - Event shape
  - Likelihood/Fisher analysis
- People tend to use standard packages

# Human resources

- KEKB computer system+Network
  - Supported by the computer center (1 researcher, 6~7 system engineers+1 hardware eng., 2~3 operators)
- PC farms and Tape handling
  - 2 Belle support staffs (they help productions as well)
- DST/MC production management
  - 2 KEK/Belle researchers, 1 pos-doc or student at a time from collaborating institutions
- Library/Constants database
  - 2 KEK/Belle researchers + sub detector groups

# Networks

- KEKB computer system
  - internal NFS network
  - user network
  - inter compute server network
  - DMZ and a firewall
- KEK LAN, WAN, Firewall, DMZ, Web servers
- Special network to a few remote institutions
  - Hope to share KEKB comp. disk servers with remote institutions via NFS
- Belle DHCP, TV conf. , Wireless LAN…

# Data transfer to universities

- A firewall and login servers make the data transfer miserable (100Mbps max.)
- DAT tapes to copy compressed hadron files and MC generated by outside institutions
- Dedicated GbE network to a few institutions are now being added
- Total 10Gbit to/from KEK being added
- Slow network to most of collaborators

# Short term plans (Summer '02)

- Software updates by the end of March
  - Better low momentum tracking
  - Less systematic errors (tracking)
  - Finer ECL calibration
- Generic run dependent MC as we take data
  - Run dependent signal MC production ?
- Reprocess all data starting from April for the summer
  - More physics skim during the DST production
- Standardize more physics tools/skims

# Long term plans

- More CPU for DST/MC production
- Faster turn around
- Distributed analysis (with local data disks)
- Better constants management
- More man power on reconstruction software and everything else
  - Reduce systematic errors, better efficiencies