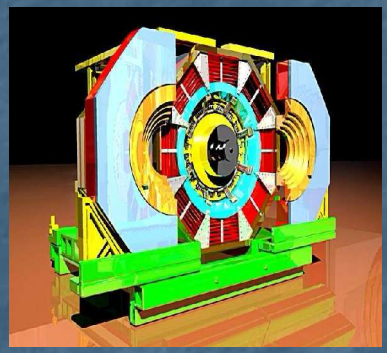
Report from e^+e^- Experiments



BES-III

by H. Neal (SLAC/BaBar) for the

BaBar, Belle, BES-III and CLEO

Experiments at

the Data Preservation and Long Term Analysis Workshop

28 January 2009

Summary of Experimental Status and Data

	BaBar	Belle	BES-III	CLEO
End of Data Taking	07/04/08	~2010	~2017	01/04/08
Collaboration end date	end of 2012	end of 2012	2017-2022	
Type of data to be preserved	raw + sim/recon (ROOT)	raw + MDST	raw+DST (ROOT)	OBJY/PDS (too dificult) preserve analysis data
quantity	2 Pbytes	~4 Pbytes	~6 Pbytes	
desired longevity of long term analysis	unlimited	5 years (until super KEKB)	15 years	superseded by B-Factories and BES-III
Simulation	Geant4	Geant 3	Geant4	Geant3
Platform	RHEL3,4,5		SLC4	
code	C++	Fortran	C++	Fortran, C, C++

Common Data Format

- favor the creation of a general data definition format
  - difference with respect to hadron collider experiments is in the needed precision of event quantities like the vertex position, boost, particle ID, etc...
  - given the lower complexity of the events, more information can be preserved
  - does not completely exclude a definition common beyond just the ee experiments
  - propose a general data definition library (DDL)
-

Common Data Definition

- suggest using complex precision analysis as a goal for creating a viable definition
- handling particle ID is a concern
- handling miss-modelling a concern
- see the possible BaBar-Belle collaboration as a test bed for developing a common data definition
  - the common data definition should allow easier combination of the data including better clarity of the quantities used in the experiments

Prototypes Exist

learn from the analysts ...

they've been archiving the data for years in the detail that is needed and with the simplicity that makes it understandable for them

In OPAL we had a few n-tuple formats that were then adapted to many different classes of analyses.

Simulation

- continued simulation ability a must
  - must be able to simulate new signals
  - validation is a concern
- need to simplify so that introducing new models does not imply touching the simulation/reconstruction code
  - assume that the simulation/reconstruction code will be too complex for any non expert collaborator to touch and still have confidence in the result. Any change may create incompatibility with the frozen reconstructed detector data.

Simulation

some projects are already down to just a couple of people that deeply understand the machinery of the code, and they are already becoming involved in future projects.

Simulation

- need standard format for generator output that would be the input to the sim/recon application
  - use any generator you wish, in whatever language you wish and on whatever platform you wish ... just output the four-vectors in a format which we should define
  - output to common data format

Simulation

- consider other means of simulation:
  - library of full detector response given initial particle track parameters before any detector interaction (Knuteson)
    - easier to maintain and ensure reproducible results
    - perhaps too late to start developing for experiments that have stopped taking data

Virtualization

- virtualization particular value is in keeping the simulation and data description (documentation and book keeping databases) alive
  - It is not a circumvention of the simplification and use of common data definition format -
  - it is but the last step after migration to a modern platform and simplification of the entire analysis/production framework

- BaBar revalidates its code for every new platform
  - (CPU, kernel security patches, 64 bit vs 32 bit, etc...)

- concern expressed that multiple platforms must be supported to increase integrity

archival system

- the main concern is to have someone to manage the archival system
  - hardware needs are minimal
    - BaBar model: one box with many cores and a large disk or other appropriate storage
- a complete documented system is needed
  - we can already run a BaBar sim/recon/analysis job on a laptop
  - all non-runtime features also needed by analysts need to function

Needs for an Analysis now & then

- If I were to start on a BaBar analysis years from now, I would at least want:
  - bookkeeping tools - data quality details, what data to use
  - access to luminosity information, beam energies, calibrations, interruptions
  - hypernews - the sum of all knowledge (not quite but close)
  - papers + support documents

Caution

- Having code running in a box is great. I can still run ancient releases of the CMS reconstruction and produce event displays on my old laptop.

- However, code can be amazingly sensitive and its possible that virtualization might for example tickle a problem like what we saw at ccin2p3.

> *BaBar simulation at ccin2p3 was stopped for many months because at a given moment the results from the site no longer could be validated (they disagreed with those from SLAC and other sites). A subtle peculiarity of their setup which actually improved the precision was the culprit. It came from an unlikely source and was found finally by computing professional that is now unfortunately retired.*

Then what ... who would work on the problem? The effect might be very small on the physics but a simple validation test might indicate that the simulation was no longer working.

Review and Approval

- very important to have a review & approval process beyond the end of the collaboration
- documentation .... documentation ... documentation
  - must have non-experts expose our documentation failings