Persistence and Data Handling Models Summary

David Brown, LBNL

SuperB Computing Workshop
12 March 2010

Background

- SuperB is a 2nd generation experiment

  - A reference computing model exists

  - broad understanding of what's needed exists

  - 100X BaBar data, ~1/2 manpower

- No totally new ground is being broken

  - multi-core as possible exception

- LHC has evolved 5-10 years beyond BaBar

  - much larger manpower pools

- Our job is facilitate the Computing TDR

What is needed for SB Computing TDR?

- Research
  - substantially new ideas needing prototyping
- Development
  - existing ideas needing refinement
- Survey
  - cost/benefit evaluation of existing options
- (Outline of) Work
  - Planning for actually getting things done

Persistence/data R&D issues

- Persistence technology (technologies?)
- Interface with transients
- dynamic conversion?
  - smart pointers
- Interface with detector(online)
- Direct access?
- Files structure
- Processing model
- Interface with database (bookkeeping)
- Interface with framework

From Tuesday Intro

Data Model Research

- Can we extend data processing model further into analysis?

  - Central production of analysis data

    - tuples, ascii, ...

  - Bookkeeping for analysis datasets

  - user-friendly interfaces (gui)
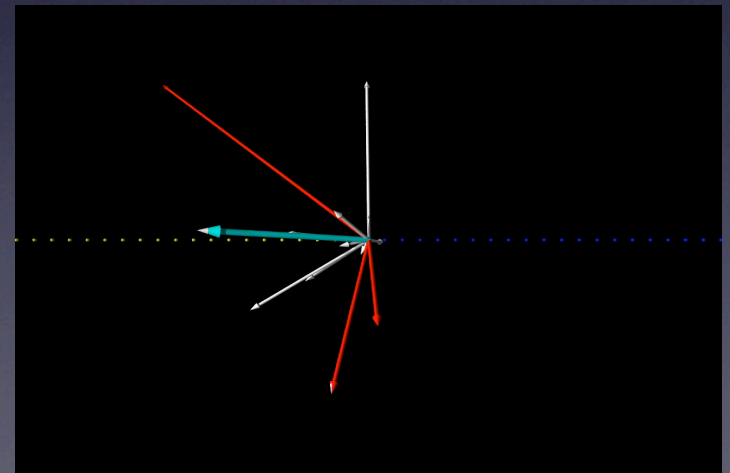
Framework Survey/Development

- Evaluation of Gaudi, CMSWS, BaBar

- requirements + problems to avoid (from BaBar)

- algorithm migration costs

- use of existing services

- external dependencies & technology bindings

- collaboration with other experiments

- compatibility with future developments (multi-core)

Persistence Survey/Development

- technologies (ROOT, hdf5, SciDB)
  - applicability to SuperB use cases
  - event store, conditions, bookkeeping
- Dictionary vs code-based schema
- standardization of analysis formats
  - extension of TupleDump?
- Compatibility with physical computing model
  - distribution across grid, tiers, clouds
  - service access from worker nodes?
- Persistent references
  - multi-write?

Persistence Research

- Forwards-compatibility & self-documentation of data format

- Collection bookkeeping

  - Provenance

  - dataset selection

- binding of conditions to datasets

  - conditions vs collection metadata

Miscelaneous

- Computing organization
  - definition of roles
  - relationship of computing to detector, analysis
  - integration of training
- use of comp. sci. undergrads?