



Kanga

Tim Adye

Rutherford Appleton Laboratory

Computing Plenary

BaBar Collaboration Meeting

10th July 2002

Overview

- Kanga Production

Ralph Mueller-Pfefferkorn, Kate Mackay,
Simon George, Dieter Best

- Recent Changes

- Moving Kanga analysis from SLAC to RAL
- ROOT3
- skimData

Nicole Chevalier

Alessandra Forti

- Kanga Development

- Pointer skims

Fabrizio Salvatore, Nicole Chevalier,
Tilmann Colberg, Ulrik Egede, Tim Adye

Kanga Production

- This is the **current status** of Objy->Kanga conversion
 - All these Kanga files (series-10 processings) are on disk at SLAC (for now) and at RAL.
- Original tag
 - AllEvents + 19 data streams: all complete
 - SPKanga: up to date with last sweep into simuboot (19 July)
 - Generic MC streams: not done
- New tag (“reskim”)
 - AllEvents data should be done in about a week
 - Nearly all 2000-1 data available now
 - 15 data streams: to be done next
 - Generic MC AllEvents + 15 streams: hopefully ~2 weeks
 - 1106 runs available so far

Moving Kanga analysis to RAL

- Since 1 July, RAL Tier A has **all Kanga** data+MC
 - RAL has kept the **majority** since Dec 99!
 - Newly converted Kanga **imported** from SLAC within **1-2 days**
 - RAL dedicated to Kanga
 - **and other non-Objectivity analysis**
 - Currently lots of free CPU and disk
- See Manny's slides from Monday
 - "RAL Tier A", **E. Olaiya**

Removing Kanga analysis from SLAC

- Older (series-8) MC+data removed from SLAC disk in May+June
 - First archived in HPSS
 - RAL copy checked against SLAC (`cksum`)
 - So far no errors found ☺
- Rest will be removed from SLAC disk at the end of July
 - May have to remove some beforehand to allow continued production
 - Keep an eye on Kanga HN just in case

ROOT3

- Migrated Kanga I/O from **ROOT 2** to **3**
 - Needed for **secure** access to ROOT daemon
 - Also allows us to **upgrade** to a recent version for other uses
 - eg. fitting, user analysis
- **ROOT 3.02-07** used in releases **>=12.0.0**
 - New version can read old Kanga files

skimData

- skimData moved out of releases $\geq 11.12.1$
 - Now called `skimDataRel` in release and symlinked from `$BFROOT/bin/skimData` (in user `$PATH`)
 - Will allow **latest version** to be used regardless of user's release
 - Other sites should take care of this with MySQL installation
- skimData now defines **conditions file** according to dataset selected
 - Generated `tcl` file includes `RootConditionsFile` and `FixedFieldStrength` specifications
- (skimData can now be used for **Objectivity** too!)
 - Allows similar selections for Objy and Kanga analyses)

Pointer Skims

- RAL currently has to store 19 streams as well as AllEvents
 - Streams allow faster access to their component skims
 - Tier C sites need only import the streams they want
 - But, entails a factor 3 overhead in disk and network
 - Original-tag series-10 data streams use 2 x AllEvents on disk
 - Fortunately size per event is still smaller than Objy
- Idea is to keep just AllEvents and generate “pointer collections” for each skim
 - Pointer files use a negligible amount of disk space
 - Could be generated at RAL from AllEvents
 - Analyses can read only their favourite skims
 - Don't have to skip other events in the stream
 - Tier C sites only need import the required skims

Pointer Skim Production

- Kanga **I/O modules** and **skimData** already understand pointer files
 - Pointer skim production now part of **standard output module**
- Working on putting this into **production**

Exporting Pointer Skims

- Pointer skims depend on **AllEventsKanga**, which is not exported to **Tier C** sites
- The plan is to **convert** from pointer to event-data files as part of **export** procedure
 - **Standalone conversion program now working**
 - ROOT application does not use BaBar Framework
 - Therefore **faster**
 - **On-the-fly conversion will be incorporated into skimImport**
 - Controlled from **Tier C** site like current stream imports
 - **ssh** used to run conversion program at **RAL**
 - Output file written to **Tier C disk** via ROOT daemon.

Efficiency

- **Pointer skim** reading is **less efficient** (per event) than reading all events, especially for sparse skims.
 - Effect probably **negligible** compared to physics code
 - Very **sparse skims** might be stored as event-data files at RAL as well
 - Studying different configurations to optimise **disk vs read rate**

Summary

- All **original-tag** data available in Kanga format
- Some **reskimmed** Kanga data available
 - Remainder in next few weeks
- Kanga analysis must move to RAL **this month**
 - All Kanga data at SLAC is also available at RAL
 - Plus some no longer available at SLAC 😊
 - RAL is ready and waiting for your jobs!
- Migration to **ROOT3** is done
- **skimData** improvements
- Developing use of **pointer skims**