

Configurable Skims Production

Nicole Chevalier
RAL Kanga Workshop
14th Jan 2003

Introduction

- Part of the New Analysis Model
 - New configurable centrally produced skims
 - To replace large AWG ntuples
 - Run on mini
 - New production every 3 months – but not all skims run.
- Test pass April – limited in both skims and functionality
- Start production ~ October
- This talk has lots of question marks!?!?

Configurable?

- Possibilities for configuration:
 - Deep copy of mini
 - Subset of mini
 - Deep copy of new micro
 - Pointers to micro
 - Subset of micro (maybe not in central prod?)
 - Any of the above + additional user data.

Additional User Data

- Quantities per event or per candidate
 - Floats, ints or vectors of floats/ints.
 - Anything more complicated?
- Composite candidates
 - What format? BtaCandidates, just genealogy, genealogy plus summary (p4, vertex).
 - Configurable?
 - Additional information?
- B candidates
 - Store lots of variables associated with these?
 - dE, mES, tagging, continuum suppression variables ...

Can do Apr Can't do, yet

- Deep copy of mini - Objy No Mini Kanga, October?
- Subset of mini – lose unwanted lists, hits, track fits –more?
- Deep copy of micro – Eric's new micro kanga format – New new micro in October? Reduced mini?
- Pointers to micro – Not until we run on Kanga
- Subset of micro - Need new new micro

Can do Apr Can't do, yet

- Additional user quantities per event or per candidate
 - Micro – needs work, mini?, pointer skims?
- Additional user composite candidates
 - Objy mini – almost sorted
 - Micro – Eric's kanga can store just genealogy – needs interface.
 - Configurable ?
 - Additional information???

Production Executable

- Run on mini (objy initially, hopefully kanga in Oct)
- Path for each skim:
 - Tag bit filter
 - Read in micro - optional
 - Micro Filter also calculates user specific quantities – optional (Do we need to read micro for pointer skims?)
 - Read in mini – optional
 - Mini filter, also calculates user specific quantities - optional
 - Write output: mini, sub-mini, micro, micro-pointer or submicro plus additions.
 - Write new tag bit of skim? (Make new AllEvents?)

AWGs Provide

- Name of skim
- Names of tag bits to filter on.
- Info about what to read in and write out, whether there is micro filter, mini filter, user additions.
 - How much config, how much tcl? – see later
- Micro/mini filter modules with any calculations of user specific quantities.

How AWGs Do User Additions

- Create lists of composite candidates.
- Evaluate other variables in filter modules and add to candidates or events
 - (Example interface (Aaron Roodman))
 - myCandidate->addToIt(“skimName_var”, skimName_var)
 - myEventTag->addToIt(“skimName_var”, skimName_var)
 - Use name convention to avoid clashes?
- In tcl, set to write wanted user additions:
 - storeAttribute set skimName_var
 - BtaCandidateList set compositeListName

Bookkeeping

- Skimming bookkeeping needs to deal with:
 - Tag, micro, pointer->micro, mini, pointer->mini, reduced mini, user defined quantities, one skim to many runs, reprocessing, reskimming ...
- In config:
 - Skim Name
 - Output (-micro, -mini, -reducedmini, -skim ??)
 - Naming conventions need defining
 - Tag bit filter ?
 - Micro filter name ?
 - Mini filter name ?
 - What user additions ? Yuck!
 - How is mini reduced? Also yuck!
 - OR reference to tcl containing filters, user addition, mini reduces
 - OR Tcl files associated by naming convention (and release)

Disk Space Requirements - Input

- Need disk space to stage mini
 - Depends on the length of skimming process?
 - 1 month or 3 months?
 - Part of disk space generally required for mini staging.
 - Size of mini: 12TB for data, 26TB MC
 - When data is staged in rotation coincide with that?

Disk Space Requirements -Output

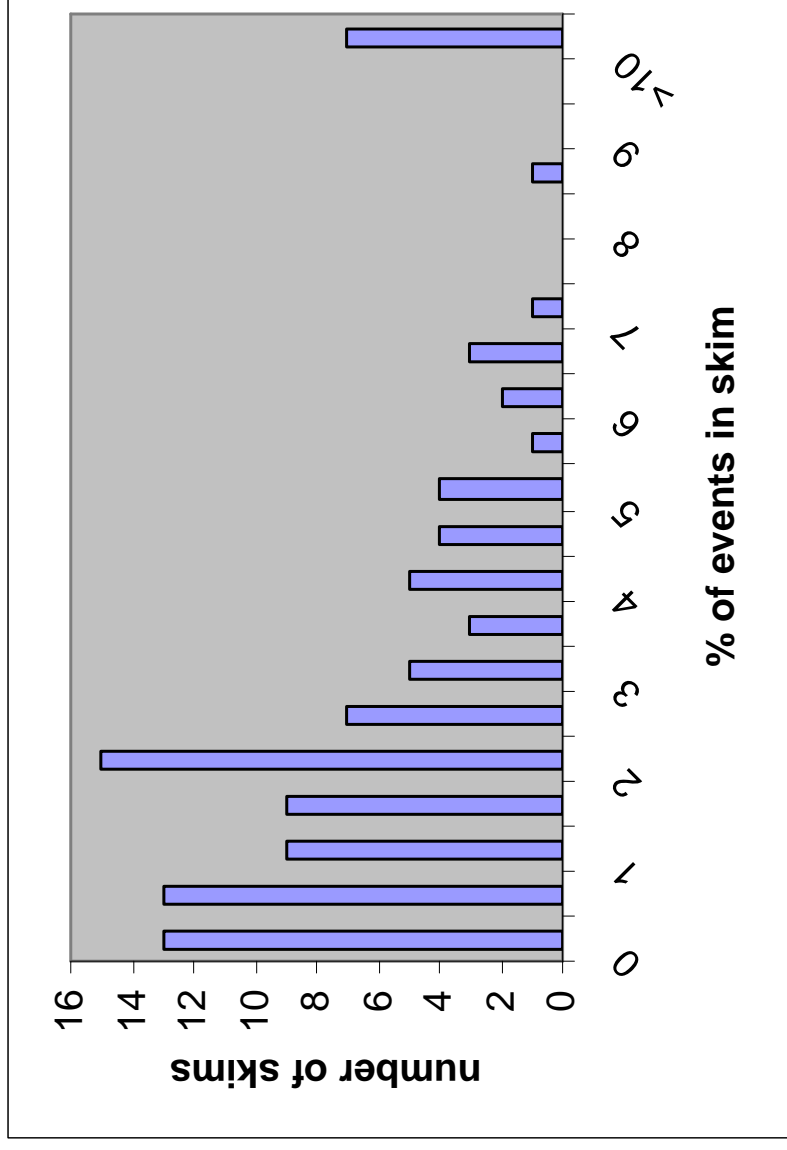
- Size of events in skims? (kB)

	AllEvents	MC
Mini	7.6	22
Reduced Mini	1.8	3.2 (no tru)
Current Micro	1.7	6.5
Eric's Micro	1.4	5.2
Reduced micro	~1	~2
Pointer skims	0	0
Additional info	~0.5	~0.5

- What we allow the AWGs – how much in deep copies, how much additional user data - depends on diskpace

Disk Space Requirements -Output

- Assume 100fb^{-1} , generic MC 300fb^{-1}
 - AllEvents Micro = 10TB (data + MC generic)
- Skim prod run quarterly, but only $\sim 1/4$ of skims run each time.



Use current
skims as
model.

Disk Space Requirements -Output

- 2 scenarios for skim outputs:
 - Skim<1%: deep copy mini, skim<2% deep copy micro sized data, else pointer to micro and user data
 - All skims = 3kB/ev data, 6kB/ev MC,
 - for 1 year 15TB
 - Skim<2%: deep copy mini, skim<5% deep copy micro sized data, else pointer to micro and user data
 - All skims = 6kB/ev data, 16kB/ev MC,
 - for 1 year 35TB

CPU usage

- CPU used:
 - Reading mini - small
 - Writing out skims - small
 - Filter calculations - Main time consumer
 - Other user specific calculations - unknown
- Estimate time from current objy skims (assuming $\frac{1}{4}$ of skims run)
 - 520 SLAC CPU days for data
 - 430 SLAC CPU days for MC
 - With 40 processors = ~1month (no hiccups).