



TM and © Laurent de Brunhoff

Status and Plans for Kanga Exports

Tim Adye

Rutherford Appleton Laboratory

Data Distribution Parallel Session

BaBar Collaboration Meeting

26th June 2001

- Kanga import tools
- Kanga statistics
- Kanga plans

Kanga Import Tools

- Tools in “production” use since February
- Documentation available
 - BaBar -> Computing -> Data Dist -> Kanga Remote
 - http://www.slac.stanford.edu/BFROOT/www/Computing/Offline/DataDist/kanga_remote.html

Kanga Import Tools – Improvements

- Minor **improvements** since February
 - Improved **documentation**
 - skimSqlMirror **more robust**
 - skimImport **uses scp by default**
 - No setup required, but
 - less efficient than **bbftp**
 - **bbftp**
 - Support version 2
 - Installation moved to **globally accessibly** area (cf. DataDist HN)
`/afs/slac.stanford.edu/public/software/bbftp/bin`
 - skimBackup **now complete**
 - Requires interface to local tape system
 - Use **import_status** also at **SLAC**
 - Useful for when we backup Kanga files at SLAC

As of 25 June.
Some skim releases combined.

Kanga Data at SLAC

Skim Release	Stream	Files	Events	GBytes	Skim Release	Stream	Files	Events	GBytes
Skim880g	AllEventsKanga	3671	301768409	940.4	K865aP1	SPKanga	59277	110804500	673.7
Skim880g	Stream1Kanga	3671	25728952	75.9	Skim880gMC	Stream1Kanga	1340	3024185	17.7
Skim880g	Stream2Kanga	3671	15684636	30.4	Skim880gMC	Stream2Kanga	1340	0	0.0
Skim880g	Stream3Kanga	3671	8405724	46.9	Skim880gMC	Stream3Kanga	1340	3062910	25.0
Skim880g	Stream4Kanga	3671	8800351	48.1	Skim880gMC	Stream4Kanga	1340	3707631	29.7
Skim880g	Stream5Kanga	3671	17846310	102.5	Skim880gMC	Stream5Kanga	1340	6407115	55.0
Skim880g	Stream6Kanga	3671	55896432	304.0	Skim880gMC	Stream6Kanga	1340	23238484	186.2
Skim880g	Stream7Kanga	3671	20785574	93.2	Skim880gMC	Stream7Kanga	1340	6534561	46.3
Skim880g	Stream8Kanga	3671	9897767	47.3	Skim880gMC	Stream8Kanga	1340	3323418	24.7
Skim880g	Stream9Kanga	3671	12906058	69.2	Skim880gMC	Stream9Kanga	1340	5504771	43.6
Skim880g	Stream10Kanga	3671	8715625	36.9	Skim880gMC	Stream10Kanga	1340	2248636	16.9
Skim880g	Stream11Kanga	3671	32106681	163.6	Skim880gMC	Stream11Kanga	1340	12649879	96.0
Skim880g	Stream12Kanga	3671	32231108	160.7	Skim880gMC	Stream12Kanga	1340	12643846	93.9
Skim880g	Stream13Kanga	3671	10869249	42.3	Skim880gMC	Stream13Kanga	1340	3851969	23.8
Skim880g	Stream14Kanga	3671	9917657	46.8	Skim880gMC	Stream14Kanga	1340	4384164	29.5
Skim880g	Stream15Kanga	3671	25966605	52.2	Skim880gMC	Stream15Kanga	1340	332177	2.3
Skim880g	Stream16Kanga	3671	5359779	23.2	Skim880gMC	Stream16Kanga	1340	1569058	12.1
Skim880g	Stream17Kanga	3671	23833525	121.8	Skim880gMC	Stream17Kanga	1340	6639032	50.5
Skim880g	Stream18Kanga	3671	8209941	37.5	Skim880gMC	Stream18Kanga	1340	0	0.0
Skim880g	TauQEDalleventsKanga	3671	36969680	109.7	Skim880gMC	TauQEDalleventsKanga	1340	8649903	42.8
Skim880g	JpsitollKanga	341	523704	2.5	Skim880gMC	JpsitollKanga	1340	1826425	13.2
Skim880g	DiLeptonKanga	341	452300	2.2	Skim880gMC	DiLeptonKanga	1340	1661210	12.0
Skim880g	BPCBhabhaKanga	1784	4574388	7.0	Skim880c	BrecoBTagKanga	3055	26609706	120.6
Skim880g	BPCElectronKanga	2305	870130	1.3	Skim880c	DiLeptonKanga	3348	4181018	17.2
Skim880g	BPCKLongKanga	3270	9449	0.1	Skim880c	DstarlnuKanga	3055	15430377	76.7
Skim880g	BPCKaonKanga	1469	1305421	5.3	Skim880c	JpsitollKanga	3348	4867950	19.7
Skim880g	BPCMunKanga	1469	5346344	6.9	K8.6.2d	SPKanga	24478	34505500	192.5
Skim880g	BPCPionKanga	1469	3053648	11.8	K863cP1	SPKanga	1945	3458000	17.6
Skim880g	BPCProtonKanga	1469	466018	1.5					
					Totals				
							213983	999617890	4530.3

MC

Old data

The Problem

- Tier A/B Kanga sites (currently SLAC, RAL, CASPUR, ...?) have to store 3 ' AllEvents to allow for all Streams
- SLAC (and probably others) needs the space to store 2001 data.

Following plans are still under discussion...

Short-term Plans

- Plan to **remove** some / all **streams** from disk at SLAC
 - At SLAC, can use AllEventsKanga (or switch to Objy)
- Hope to develop Kanga “**index collections**” for all **skims**
 - Allow **efficient** access to data in AllEventsKanga files
- According to Bob’s survey, some streams are **not exported** by anyone
 - Streams 1, 3, 5, 8, 9, 10, 15, and 16
 - These streams will be **removed** from **disk** first
 - Will be **archived** in HPSS
- Will provide plenty of **warning** before removing anything!

Medium-term Plans

- As space grows **tighter** want to remove other data
 - **Older** (eg. 2000) streams removed first
 - Allow time for remote sites to copy from SLAC **first**
 - **RAL** (or other Tier A sites?) may become **disk repository** of full Kanga dataset
- We are looking into a scheme for **automatically** generating **skim** files during export to **Tier C** site
 - Could use Kanga index collection files
 - Tier C sites could export just the **skims they want**
 - Need to check that this does not place too much **CPU load** on SLAC
 - **Trade disk for CPU**
 - **skimData** can automatically adjust TCL file

Summary

- Kanga **export tools** up and **running**
- **Disk space** at SLAC and elsewhere is problem
- Need to consider **other models** than simply keeping everything on disk