Offline Monitoring Report

- Status of current launch
- Errors reported from SciComp in offline monitoring jobs
- Relating event numbers and real time
- SWIF

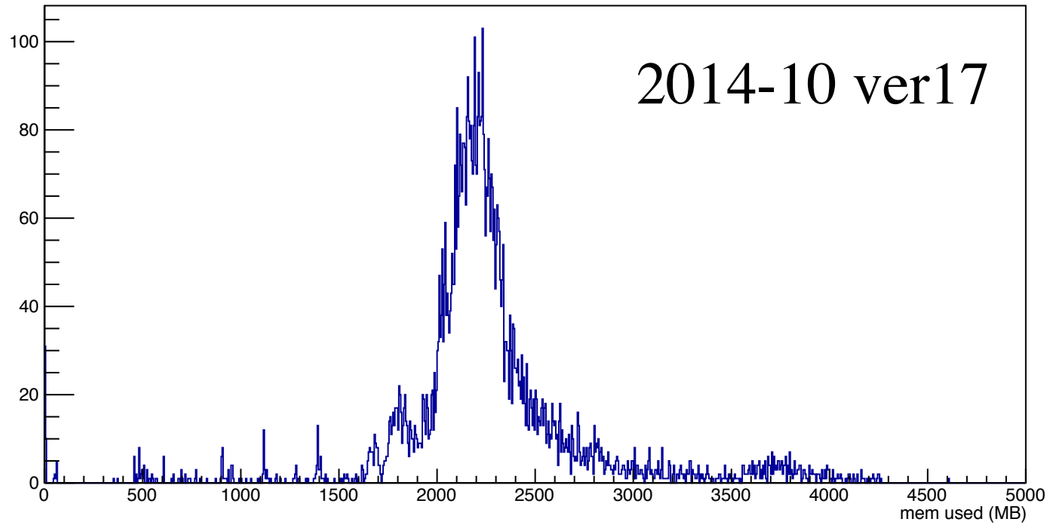
June 24, 2015
Kei Moriya

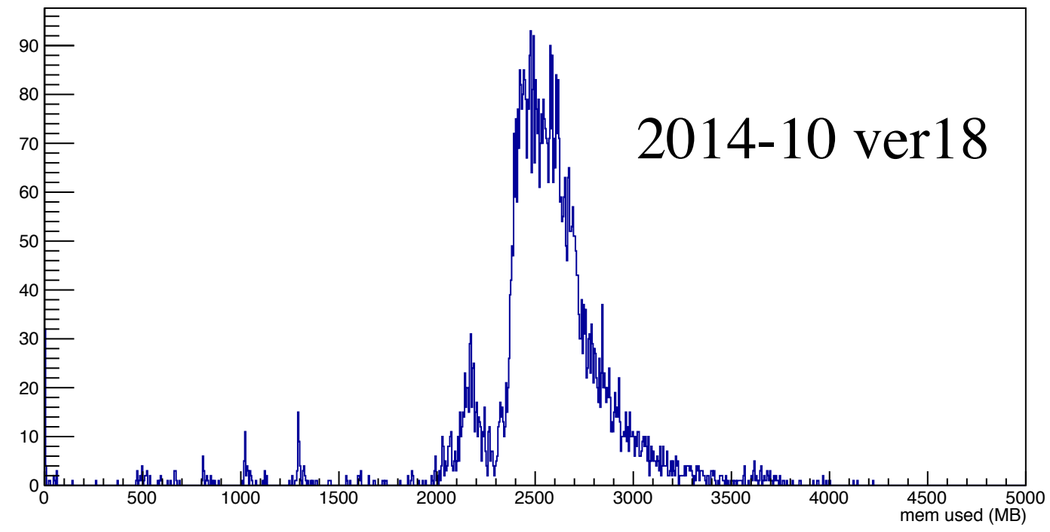
Current Launch

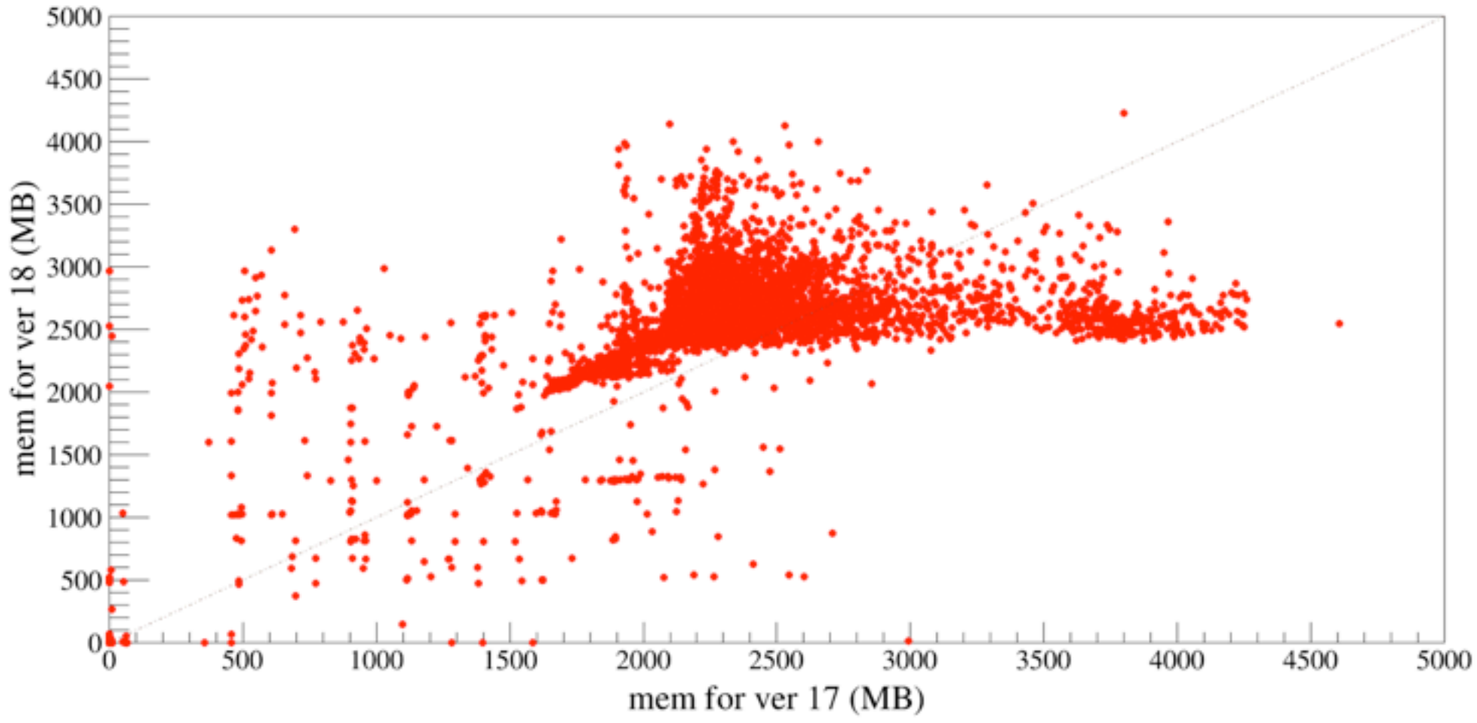
- 2014-10 ver18, 2015-03 ver08 launched on June 19 (Fri)
- 7368 jobs for 2014-10, 4298 jobs for 2015-03

Change in Memory

- Current launch uses more memory on average than previous ones







Errors for Jobs

- Checking from results of 2014-10 ver17, 2015-03 ver07 (launched June 5)
- Non-NULL errors:
  1. Job timed out.
  2. Job exceeded resource limit.
  3. No job status in batch system and we never recorded a finish.
  4. fail to get input file
  5. Job failed with unknown reason.

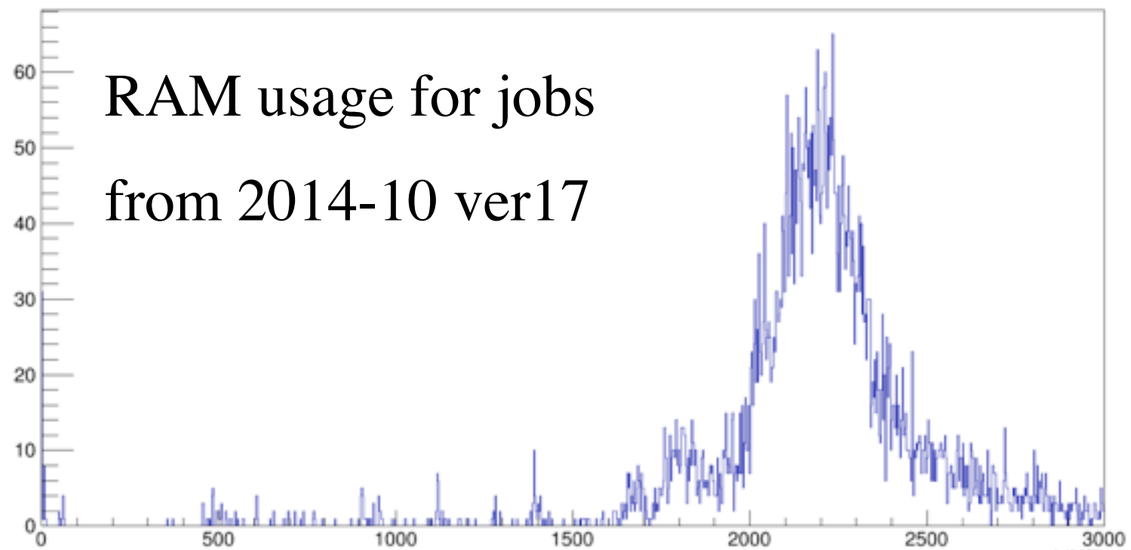
error	2014-10	2015-03
1	81	33
2	274	12
3	1	2
4	1	8
5	88	2
total	445	57

1. Job timed out.

- Typically hd_root gets stuck on single event, nothing happens until time limit reached
- Can either be from bad data file or bad software
- No log files, can’t find where job got stuck

2. Job exceeded resource limit.

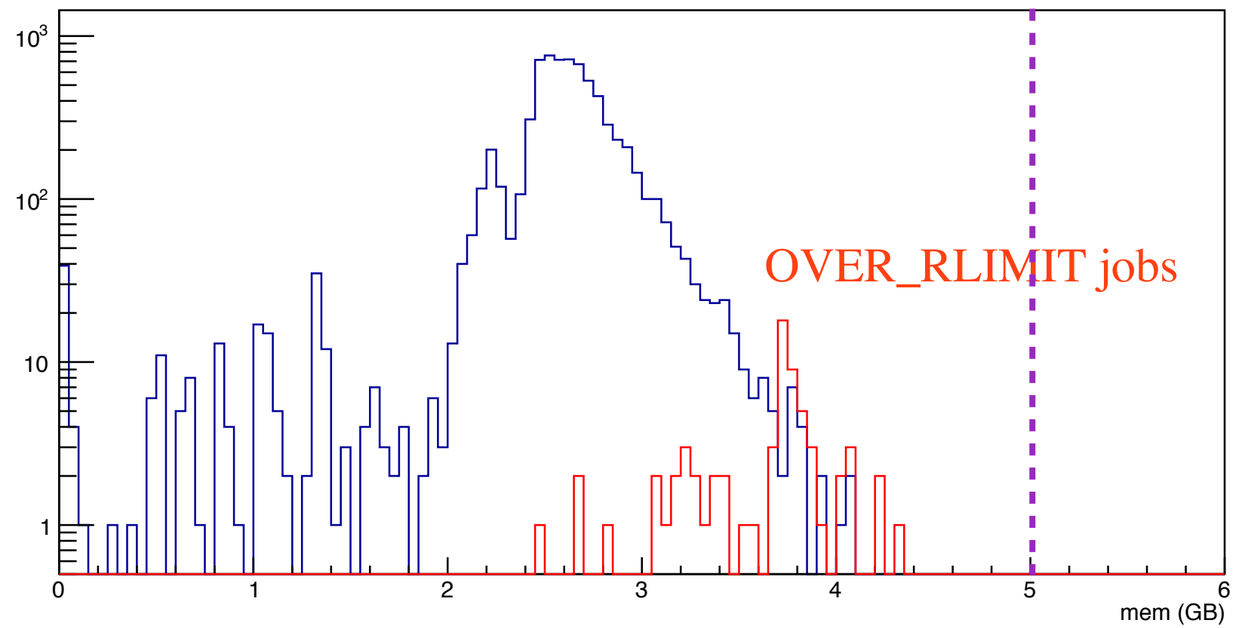
- Due to more resources needed than requested
- Asking for 40GB of disk space, 5GB of RAM



- For most failed jobs, mem usage is < 5GB, vmem usage is > 5 GB
- Happens in large runs with hundreds of files (e.g., runs 2439, 2711, 2792, 2931)
- May have to do with CDC/FDC writing out more data/event - see David L.'s talk at previous collab. meeting (will be fixed with new firmware for fADC125's?)

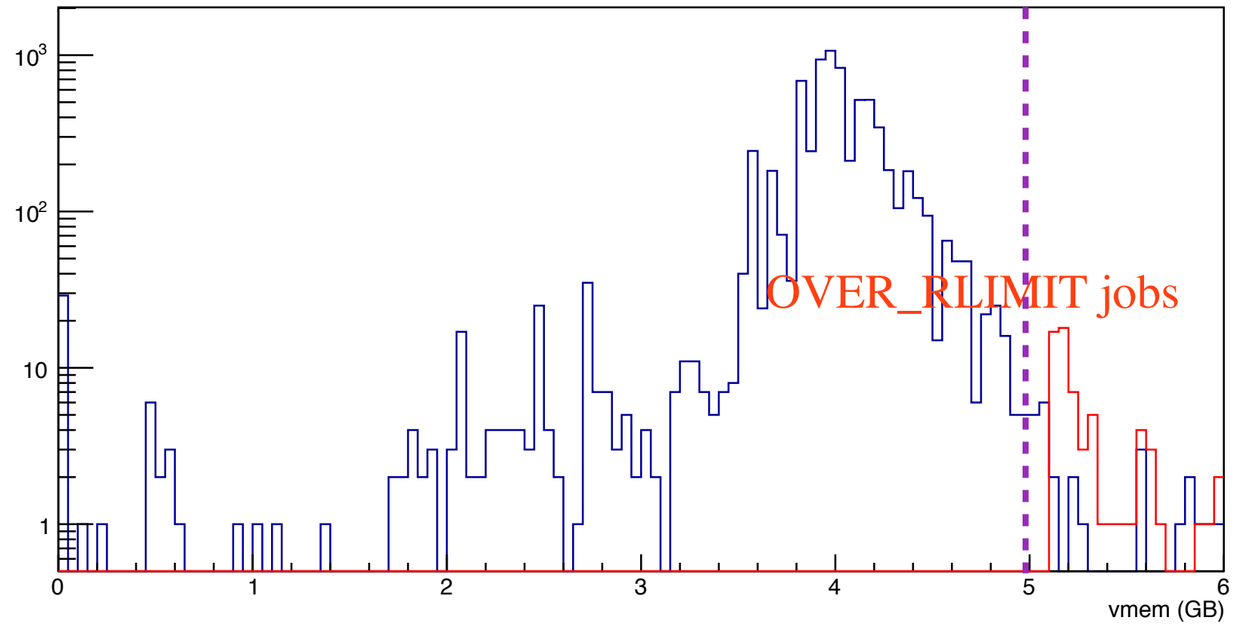
2. Job exceeded resource limit.

- Asking for 40GB of disk space, 5GB of RAM



mem usage 2014-10 ver18

vmem usage 2014-10 ver18



3. No job status in batch system and we never recorded a finish.

- Occurred 5 total times for 2014-10 ver11-ver17, 2015-03 ver02-07
- For 2014-10 ver17 and 2015-03 ver07 (launched simultaenously), all 3 failed jobs were on same host, farm12030
- The other two times we do not have a record of what the hostname was

Sent CCPR to SciComp:

```
node farm12030 had xfs errors with the
local disk and was marked offline on Monday
June 8:

[root@farmpbs14:~]# pbsnodes -ln
farm12030              down,offline
needs rebuilt, xfs errors
```

Seems like an issue with the host, would be resolved with re-submission

4. fail to get input file

- For 2014 Fall data, ~85 files consistently fail
- These files are due to files written to tape in wrong run directory, e.g. :

```
/mss/halld/RunPeriod-2014-10/rawdata/Run002134/hd_rawdata_002135_000.evio
```

- Files are found and registered in monitoring jobs as Run*/hd_rawdata_*
- Less severe for 2015 Spring data (19 files), and files have been copied to correct dirs

Do we bother to move 2014 Fall files?

Version	Total (%)	Success (%)	Over Limit (%)	Timeout (%)	Failed (%)	NULL (%)
VER09	7369 (100.00)	6883 (93.40)	265 (3.60)	111 (1.51)	0 (0.00)	110 (1.49)
VER10	7369 (100.00)	6625 (89.90)	186 (2.52)	469 (6.36)	85 (1.15)	4 (0.05)
VER11	7369 (100.00)	6066 (82.32)	232 (3.15)	982 (13.33)	85 (1.15)	4 (0.05)
VER12	7369 (100.00)	6936 (94.12)	214 (2.90)	74 (1.00)	141 (1.91)	4 (0.05)
VER13	7369 (100.00)	6974 (94.64)	234 (3.18)	72 (0.98)	85 (1.15)	4 (0.05)
VER15	7369 (100.00)	6960 (94.45)	245 (3.32)	70 (0.95)	86 (1.17)	8 (0.11)
VER16	7369 (100.00)	6883 (93.40)	338 (4.59)	59 (0.80)	85 (1.15)	4 (0.05)
VER17	7369 (100.00)	6923 (93.95)	274 (3.72)	81 (1.10)	87 (1.18)	4 (0.05)

From https://halldweb.jlab.org/data_monitoring/launch_analysis/index.html

5. Job failed with unknown reason.

- 1 file in 2014 Fall data, 8 files in 2015 Spring data
- stdout, stderr files do not exist for these jobs
- Error is rather rare, depends on launch... something on SciComp side?
- Files that had this error are not common between launches

Sent CCPR to SciComp:

```
We don't record more info that this on
the status of these nodes or errors at
that time - it is provided through the
job reporting interface. Try
submitting jobs via the swif tool --
it will give all the info it has about
the jobs, and even retry them.
```

swif may be the way forward...

2014-10

ver	times
11	0
12	56
13	0
15	1
16	0
17	1
total	58

2015-03

ver	times
2	0
3	0
4	0
5	0
6	0
7	8
total	8

Miscellaneous Items

Raw Data File Sizes

- Checked all files in 2014-10 and 2015-03, no files of size 0
- This was raised by Franz Klein during collaboration meeting
- Thought I saw this before....

Segmentation fault

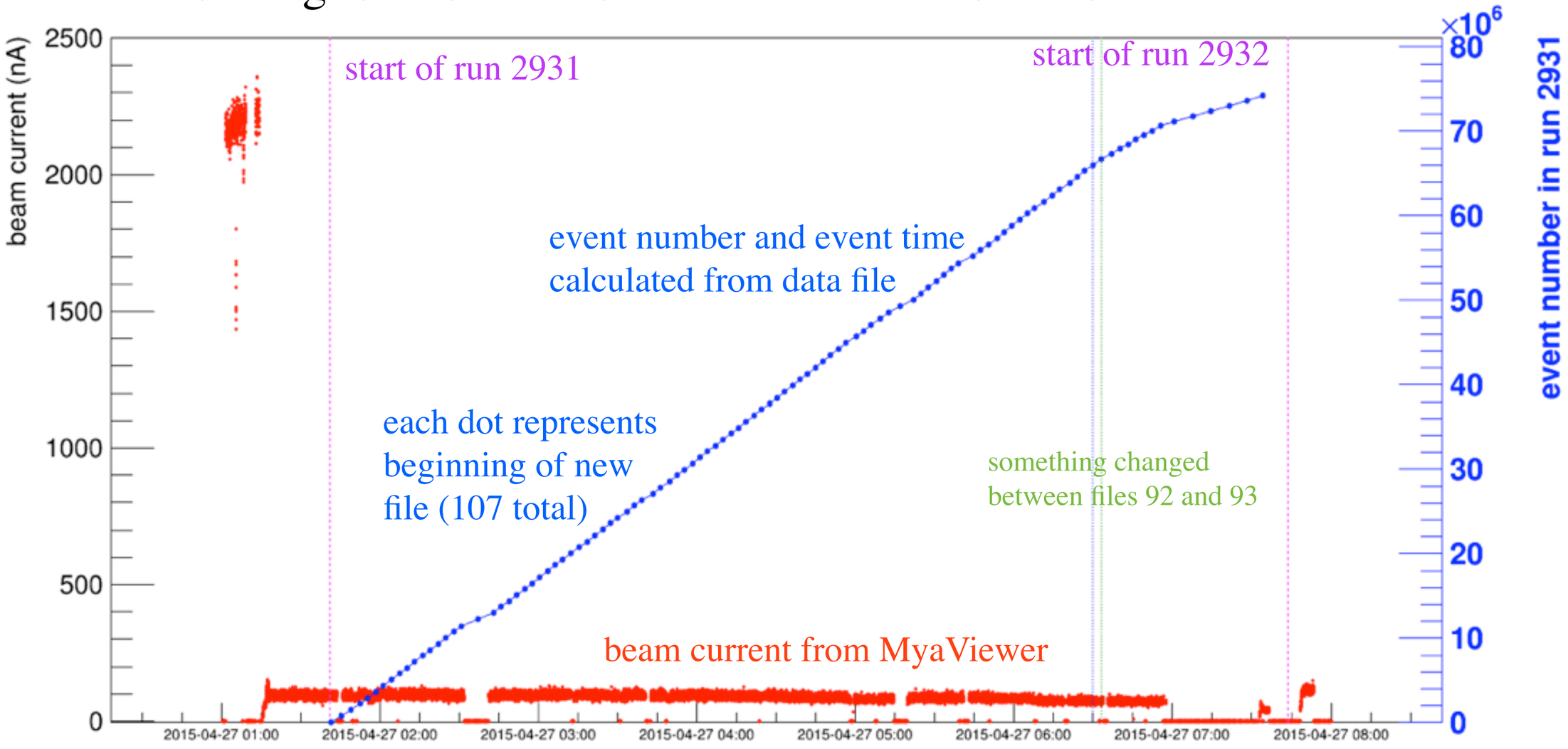
- Currently not checking for segmentation fault for launch statistics
- Could be due to problem with file or software
- 260 files in 2014-10 ver17, 27 files in 2015-03
- Will add to table of statistics

Mystery Files

- For run 3185, 35 files exist
- For 2015-03 ver07, only 25 log files, ROOT files exist
- Auger returns no error for these files (status = DONE, exitCode = 0, result = SUCCESS, error = NULL)

Relating Event Time to Real Time

- Using the event number and roctime of each event, we can calculate the event time
- May be useful for diagnosis of relation between beam conditions and data



SWIF Usage

```
> swif create my_workflow
```

create workflow

swif command

swif options

```
> swif add-job -workflow my_workflow -project gluex -track reconstruction -cores 6 -disk 40g
-ram 5g -time 8h -os centos65
-input hd_rawdata_002333_000.evio mss:/mss/halld/RunPeriod-2014-10/rawdata/Run002333/
hd_rawdata_002333_000.evio
-stdout /home/gxproj1/logfiles/stdout
-stderr /home/gxproj1/logfiles/stderr
-name my_job
/home/gxproj1/script.sh TAGH_online 002333 000 6
```

script to execute with options

register job

```
> swif run my_workflow
```

run jobs

```
> swif status my_workflow
```

check status

```
> swif status my_workflow -runs
```

check status for individual jobs

```
> swif cancel my_workflow -delete
```

cancel jobs, option -delete removes workflow

Considering SWIF For Future Usage

- Much quicker job submission than jsub (does not use Java, instead calls Auger's RESTful interface for a whole batch of jobs)
- Chris working on implementation of job info (-report option)
- Sean working on getting calibration train going
- Still a few bugs to sort out (?)
- Switching costs from current system

Moving Forward

- Next launch is July 3 (Fri) ???