Production Processing

Adam Lyon

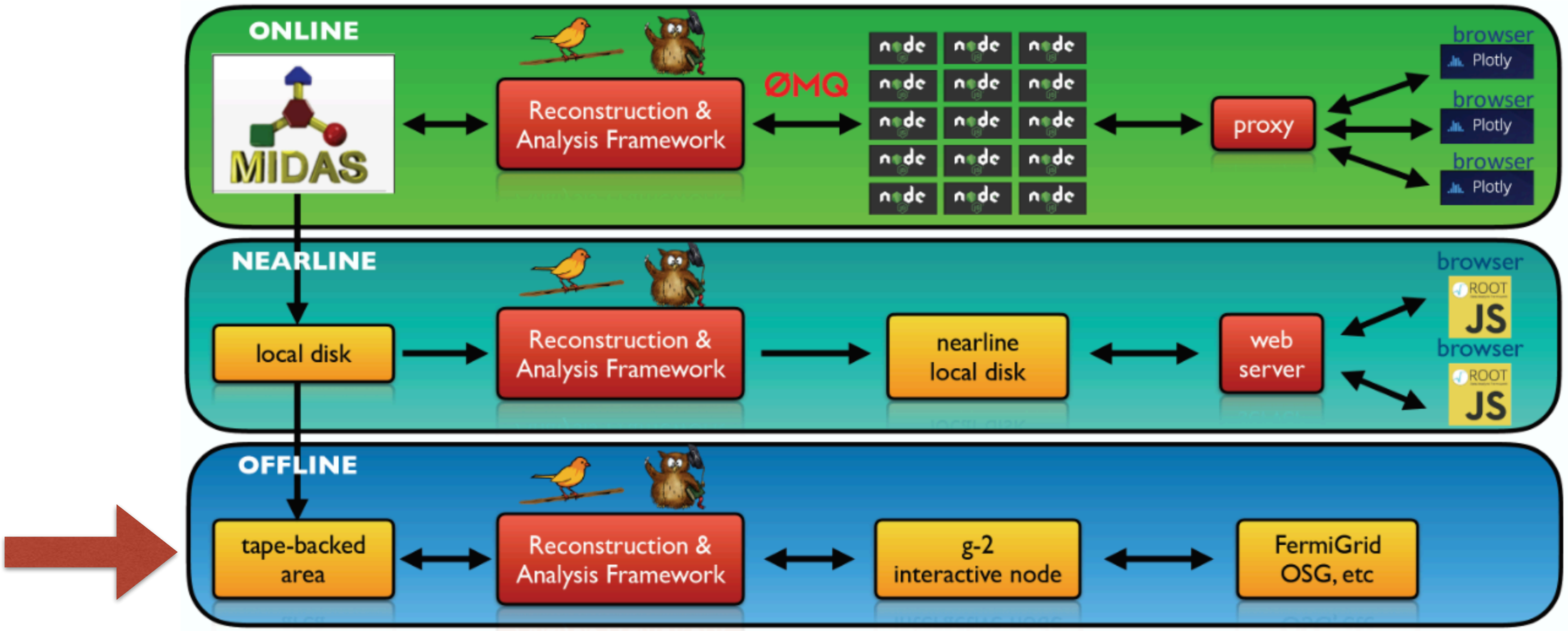
Operational Readiness Review

October 2, 2017

Charge question

4) Are there robust plans for **data processing** and data analysis? Have **adequate resources from the laboratory and the collaboration** been identified for data analysis to meet these goals?

Online/Nearline/Offline



Looks like duplication, but remember that only offline has the requirement to process ***all*** of the data … Nearline is used for monitoring & collecting calibration

Key fact: *The reconstruction code is the same for all (art framework)*

Offline Production Overview

- We have a production team
- The production framework is set up
- Production handles both Simulation and Data
- Calorimeter reconstruction works very well
- Other detector reconstruction works but integration is needed
- Some, but not all, calibrations are integrated into the production workflow

- Still production robustness work to do
- Still profiling work to do (both memory and speed)
- Still integration into reconstruction workflow to do for both code and calibrations

- The Commissioning Run exercised all of the above

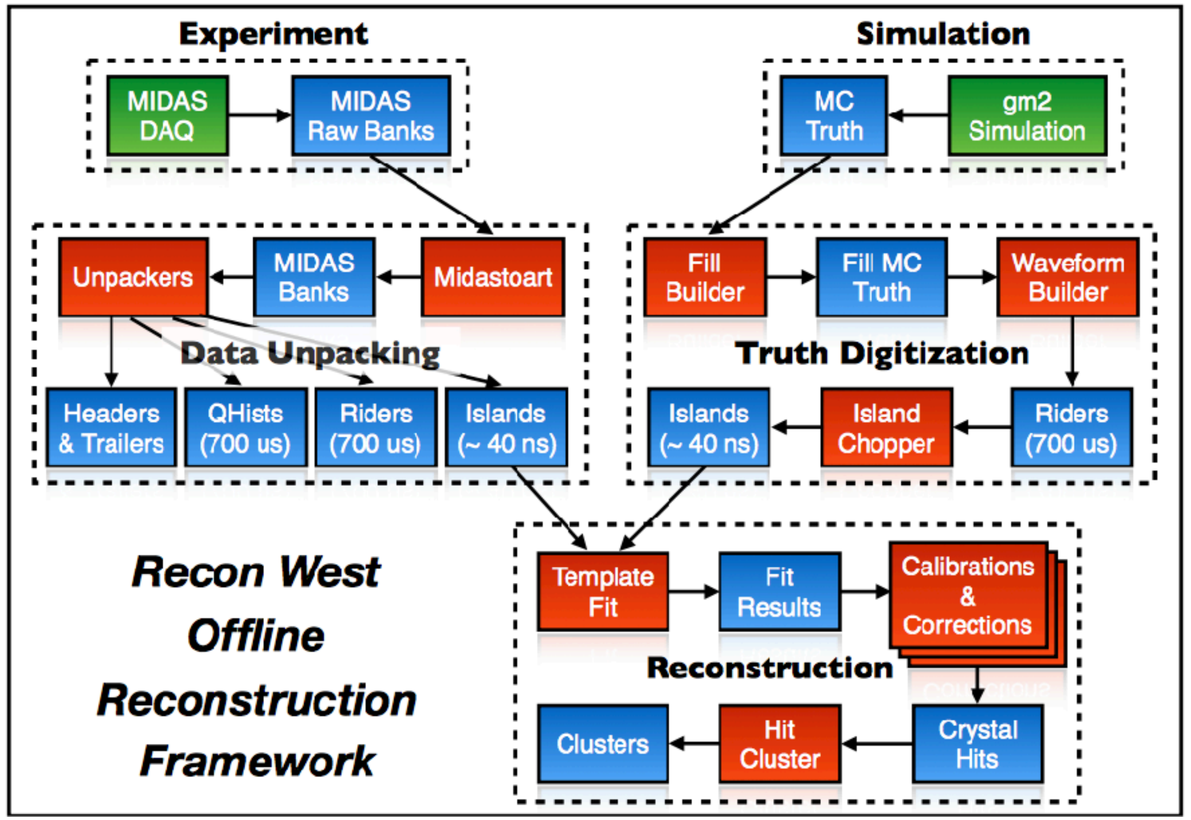
Processing

- Workflow (example on right)

- Right now we do unpacking and reconstruction in separate jobs
  - Unpacking requires large memory - fixing
  - Separating tasks increases flexibility (but at the cost of extra work for the production team)

- Calibration workflow is mature for calorimetry, becoming mature for tracking, and in progress for auxiliary detectors.

- DB assistance will be important



Multithreading — we're ahead here

- 24 independent calorimeters
- Multithreading takes advantage of multiple threads/processors to
  - simultaneously unpack the calorimeter banks
  - simultaneously reconstruct for calorimeters

- Upshot:
  We use the same technology that *art* developers use
  Same code runs everywhere (automatically adapts for # of available cores)

- **Details**
  We use Intel Thread Building Blocks (TBB) library
  Compatible with the future art v3 (multithreaded art)

  TBB automatically sets the threads appropriately given the number of available cores

  Unpacking needs a lot of memory at the moment (multiple copies of banks are made - this will be fixed).
  Multithreading allows us to use the multiple processors efficiently

Commissioning Run experience

- Understanding memory usage issues

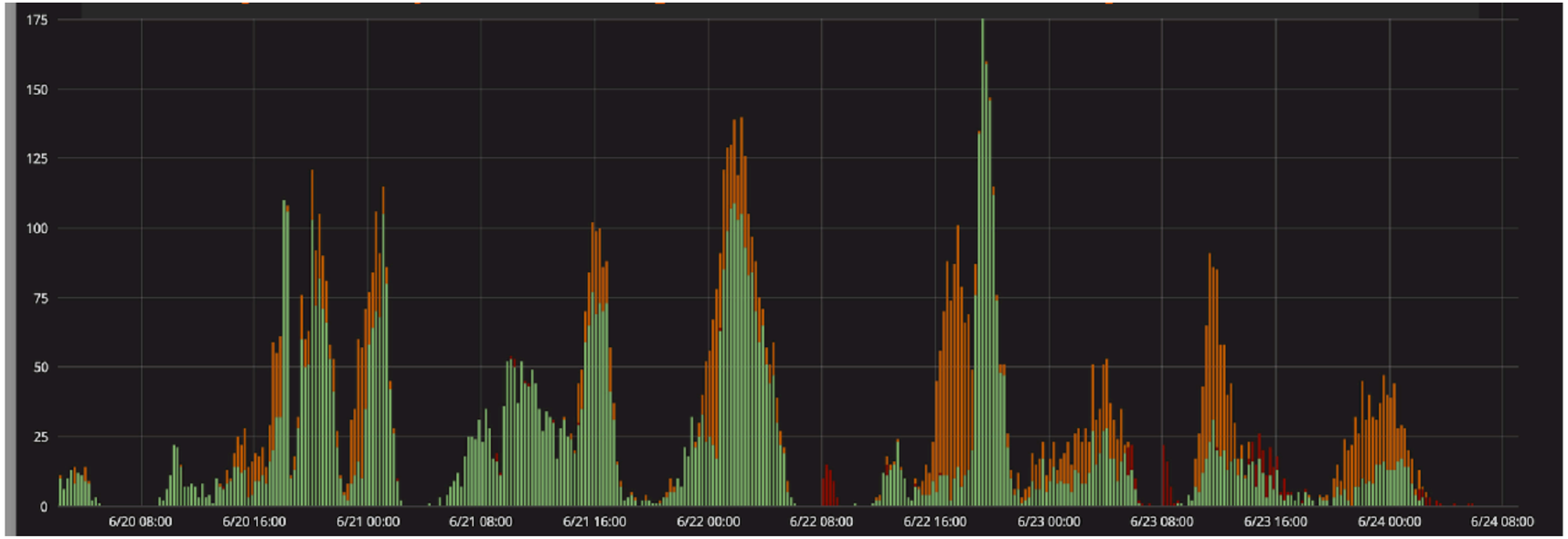
- Ran full reconstruction (but no track fitting) Multithreading!

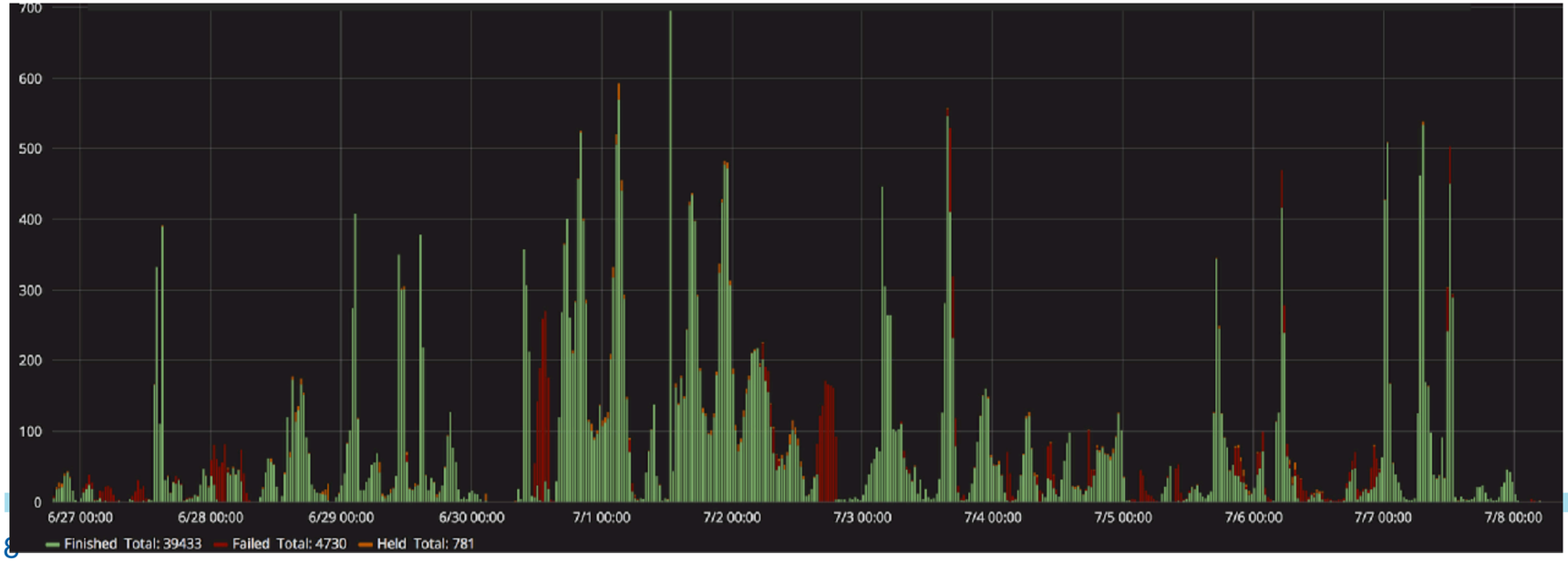
Data Tier	File Format	Data Size	Number of Files	Memory Usage
raw	midas	20 TB	15,167	-
unpacked	art	19 TB	14,933	6-8 GB
reconstruction	art	1.7 TB	14,929	4 GB

- Jobs ran daily and delay between online data capture and offline production was usually a few hours
- All 20TB of data were reconstructed in 4 days

- Large fraction of jobs were run offsite (but likely will stay onsite for FY18 data production and use OSG for simulation/ analysis)

Commissioning Run Production





Memory overages caused held jobs (yellow)

This is understood

Much better

FY18

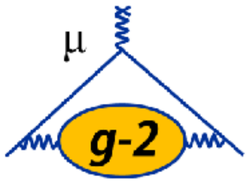
- We have month-by-month estimates rates
- An input to our requests to Scientific Computing Division

	October 2017	November 2017	December 2017	January 2018	February 2018	March 2018	April 2018	May 2018
Days	31.0	30.0	31.0	30.0	31.0	30.0	30.0	31
Data Taking Days per Month	5.0	5.0	22.0	21.3	22.0	21.3	21.3	22.0
Comment	Laser	Laser	Physics	Physics	Physics	Physics	Physics	Physics
Trigger (Fill) Rate	12.0	0.0	12.0	12.0	12.0	12.0	12.0	12.0
Events (Fills) * 10^6	0.0	0.0	22.8	22.1	22.8	22.1	22.1	22.8
Cumulative Events (Fills) * 10^6	0.0	0.0	23.0	45.1	67.9	90.0	112.1	134.9
Muons * 10^11	0.0	0.0	0.4	0.7	1.1	1.4	1.8	2.2
Cumulative Muons *10^11	0.0	0.0	0.4	1.1	2.2	3.6	5.4	7.6
Detected Positrons * 10^9	0.0	0.0	24.6	23.9	24.6	23.9	23.9	24.6
Cumulative Detected Positrons * 10^9	0.2	0.2	24.8	48.7	73.3	97.2	121.0	145.7
Incoming DAQ data [MB/s]	19400.4	0.0	19400.4	19400.4	19400.4	19400.4	19400.4	19400.4
Rate Prescaled data written to tape per [MB/s]	19.4	0.0	19.4	19.4	19.4	19.4	19.4	19.4
Rate T-method data written to tape [MB/s]	112.5	0.0	112.5	112.5	112.5	112.5	112.5	112.5
Rate Q-method data written to tape [MB/s]	121.3	0.0	121.3	121.3	121.3	121.3	121.3	121.3
Tracker data rate [MB/s	0.0	0.0	9.7	9.7	9.7	9.7	9.7	9.7
Rate other data written to tape [MB/s]	1.0	1.0	5.0	5.0	5.0	5.0	5.0	5.0
Total Fast DAQ EB data rate [MB/s]	127.1	0.5	133.9	133.9	133.9	133.9	133.9	133.9
Total Fast DAQ data written to tape [TB]	54.9	0.2	254.7	246.5	254.7	246.5	246.5	254.7
Cumulative Fast DAQ data written to tape [PB]	0.3	0.3	0.6	0.8	1.1	1.4	1.6	1.9
Field data written to tape [TB]	31.0	30.0	31.0	30.0	31.0	30.0	30.0	31.0
Cumulative field data written to tape [PB]	0.227	0.257	0.288	0.318	0.349	0.379	0.409	0.440
Total DAQ data on tape [PB]	0.6	0.6	0.9	1.2	1.5	1.7	2.0	2.3
Reco Size Per Fill [MB] (drop T islands after 3/18)			20.0	20.0	20.0	6.0	6.0	6.0
Reco Calo CPU s / fill (max)			1.0	1.0	1.0	1.0	1.0	1.0
Reco Tracking CPU s / fill (rough order of mag)			100.0	100.0	100.0	100.0	100.0	100.0
Reco Tape [TB]	54.9	0.2	254.7	246.5	254.7	246.5	246.5	254.7
Cumulative Reco Tape [PB]	0.3	0.3	0.6	0.8	1.1	1.4	1.6	1.9
Fraction of Reco on Tape-Buffer	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Reco on Tape-Buffer [TB]	11.0	0.0	50.9	49.3	50.9	49.3	49.3	50.9

Tools

- We use the standard Fermilab SCD “FIFE” tools
  - SAM, Fermi-FTS, dCache, tape, POMS, …
- We only run production using released code
- Online data is stored quickly to tape from the online disk via Fermi-FTS
- Offline processing starts with raw data on tape (or simulation)

Production Team Personnel



- 2016 Computing Review recommended the formation of a production team

- Liang Li (Shanghai) — coordinator
- Anna Driutti (Udine, Italy) — local coordinator*
- Tammy Walton (FNAL) — outgoing local coordinator — wrote common script
- Adam Lyon (FNAL - SCD Liaison) — temporary surge
- Nandita Raha, Nam Tran [New]
- Wes Gohn, Kim Siang Khaw, Renee Fatemi (Ex. officio)

No one is on production full time

*Not local at the moment

Offline Production Performance

- Running on the grid requires many moving parts run by many different people to all be working an harmony
- Robustness is challenging — understanding failing jobs is a time sink
- We need to understand the system better, understand debugging tools that are available, and work the way that SCD wants to be engaged (e.g. submitting tickets)
- SCD will need to provide help for us to understand tools and debugging and optimize our scripts, etc. POMS is also important.
- We need to initiate and coordinate. Adam has started helping here.

Code performance

- Commissioning run exposed memory usage in unpacking
- All code is in flux, for good reasons (e.g. tracking improvements)
- We need a profiling effort (memory and speed). Will ask SCD experts for help. Low hanging fruit are likely
- CPU time estimates assume fairly poor speed performance, so a large number of simultaneous jobs are needed to keep up with DAQ
- Pressure on production team will be high; any and all improvements adds breathing room

Costs

- Requests are reviewed yearly in the SCD's "SCPMT" process

- Scientific Computing Portfolio Management Team

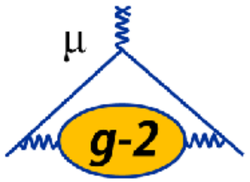
Tape	$ per TB
T10K Media	30
Migration	30
Total tape	60
CPU	$ per hour
1 Core	0.01
Disk	$ per TB
dCache	125
NAS	150
SCD People	k$ per FTE
Support	150

SCPMT FY18 Requests	**Amount**	**Units**	**Cost k$**
Data Processing CPU (onsite)	18	M core-hours	180
Simulation CPU (offsite)	9	M core-hours	0
dCache Tape Backed	400	TB	50
dCache Scratch	300	TB	38
dCache Persistent	200	TB	25
dCache Write Pool	100	TB	13
NAS Storage	60	TB	9
Tape DAQ (2 copies)	4400	TB	264
Tape Reco	1800	TB	108
Tape Simulation	1000	TB	60
Total Computing and Media			**747**
SCD Support Services	10	FTE	1,500

- Note that tape costs are more than usual due to migration need

- SCD Personnel on experiment for FY18:
  Adam (Senior-Sci 0.5 FTE) - SCD Liaison, Production surge
  James Stapleton (Post-doc 1.0 FTE) - Simulation, Release manager
  Christ Stoughton (Senior-Sci 0.5 FTE) - *Kickers (not software/computing)*

Summary



- We are starting from a good place
- We've gained good experience running production in the commissioning run and simulation
- Learned that there's work to do, especially in the area of robustness, profiling and DB. Will need SCD to help, at least in the short term
- Tape and disk needs should be adequate for FY18 based on data taking estimates