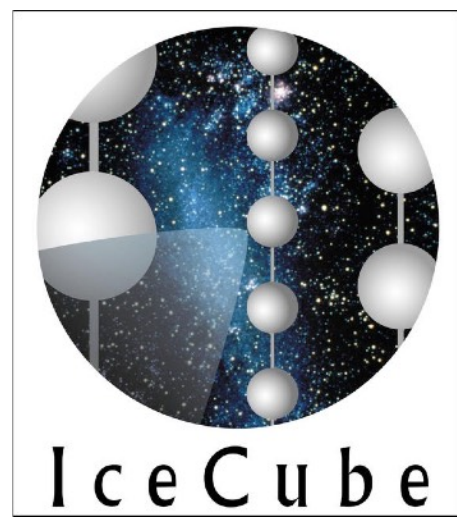


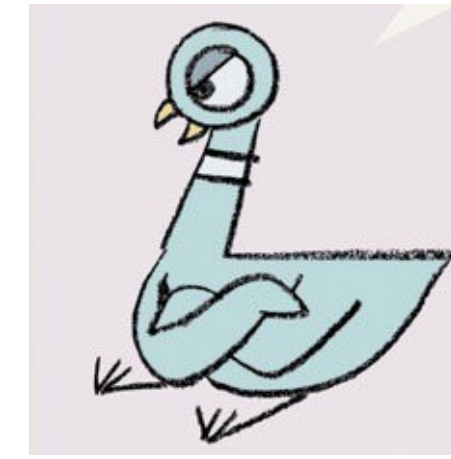
Data Filtering - Now and Future

IceCube 2023 Bootcamp

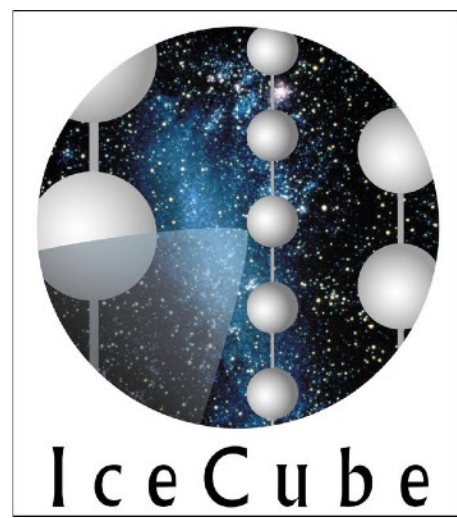
E. Blaufuss (@blaufuss) June 9, 2022



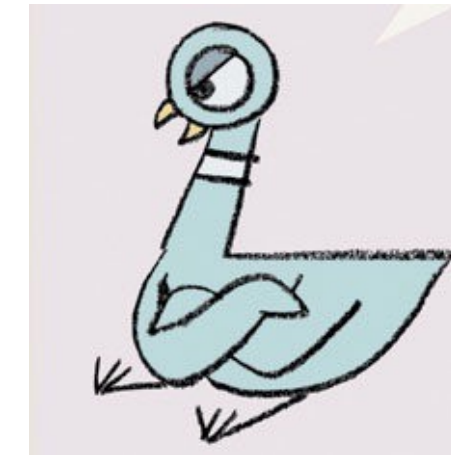
Outline



- Overview of current filtering
 - Pole
 - Level2 and beyond
- Ideas for future
 - Most events north in SuperDST

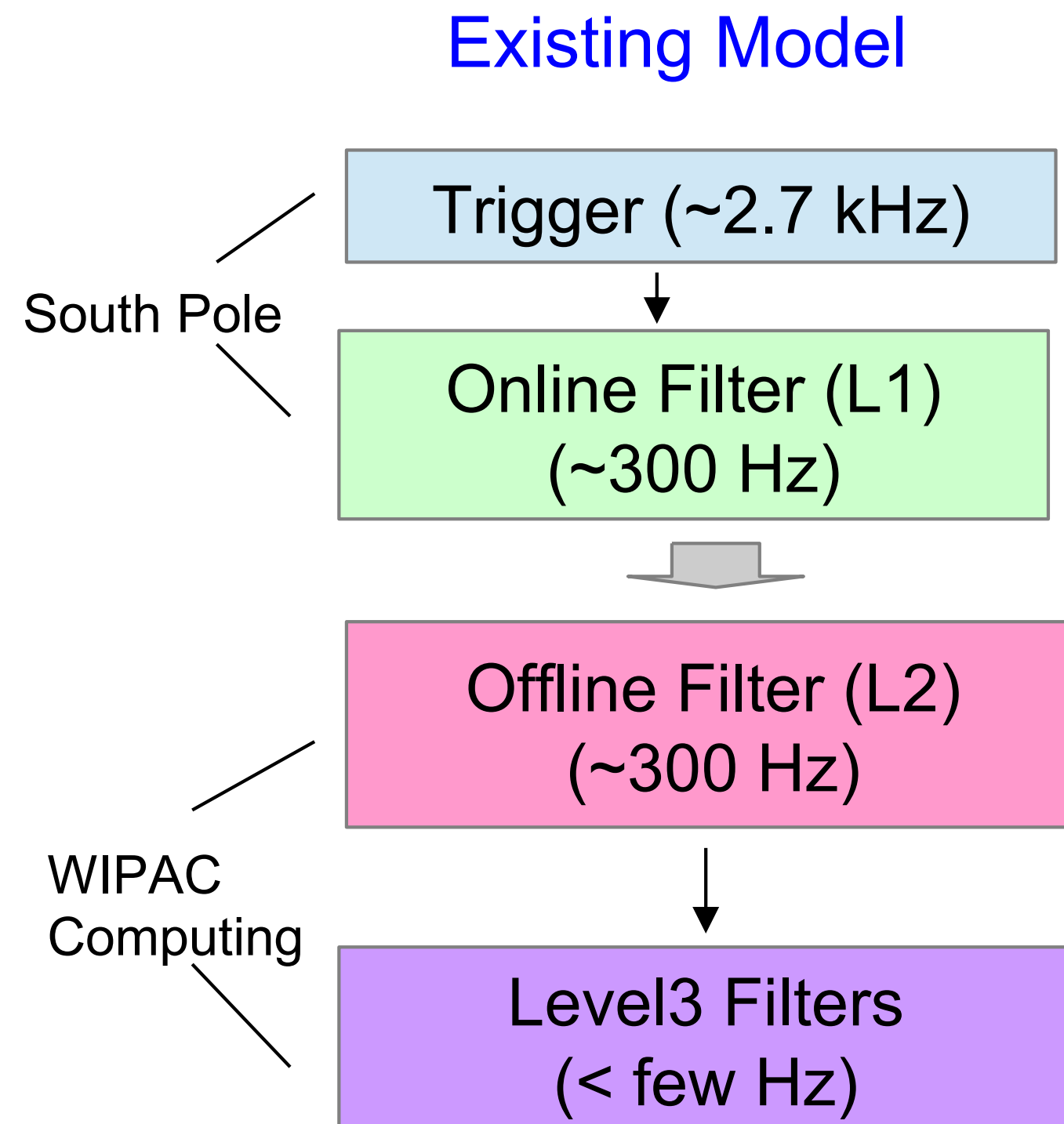


The challenges



- Remote location of IceCube makes it difficult to get IceCube data out. We had two options:
 - Save all events, and deal with them “next year”
 - Filter events in real-time, send a subset north immediately.
- We would like to start physics analysis of IceCube data as soon as possible.
 - Choose realtime filtering.
- Raw IceCube data/trigger rate is high (IC86)
 - ~2750 Hz
 - ~1000 GB/day of raw data
- Our satellite allocation (TDRS) is ~100 GB/day.
 - Need to reduce data volume by ~10%
- We’d also like to perform real time analysis, alert others in the event of interesting detections,
- Realtime analysis of data is also important for monitoring detector quality

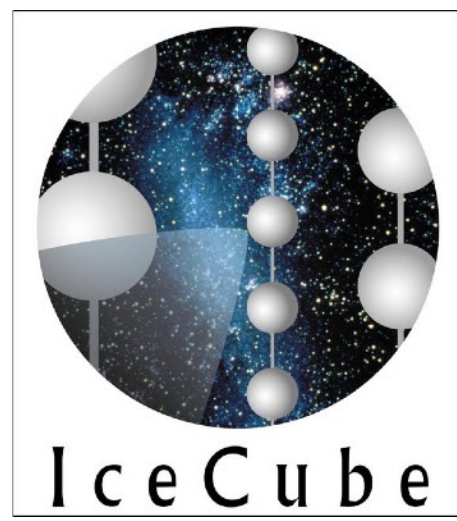
Current Filtering Overview



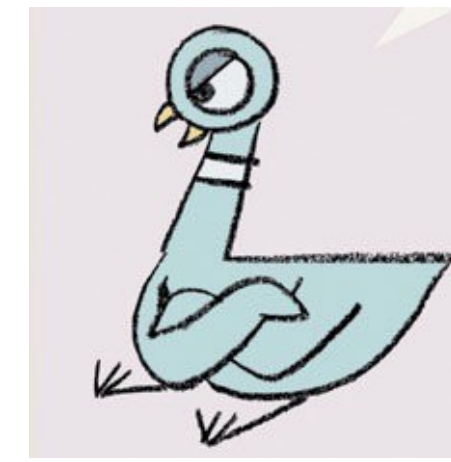
Courtesy of Naoko Kurahashi Neilson

- Storage of multiple copies of data
 - L1 (from pole) - ~75 GB/day
 - L2 (in North) - ~150 GB/day
 - Multiple copies at L3 as well, but these are more modest sizes...
 - Multiplies for MC samples
- Compute intensive.
 - Current L2 - ~1 hr/cpu per ~2 minutes of data..
 - Additionally, much repetition of processing done from L1, and again in L3+...
- Wasted effort to understand overly complex system for each new person...

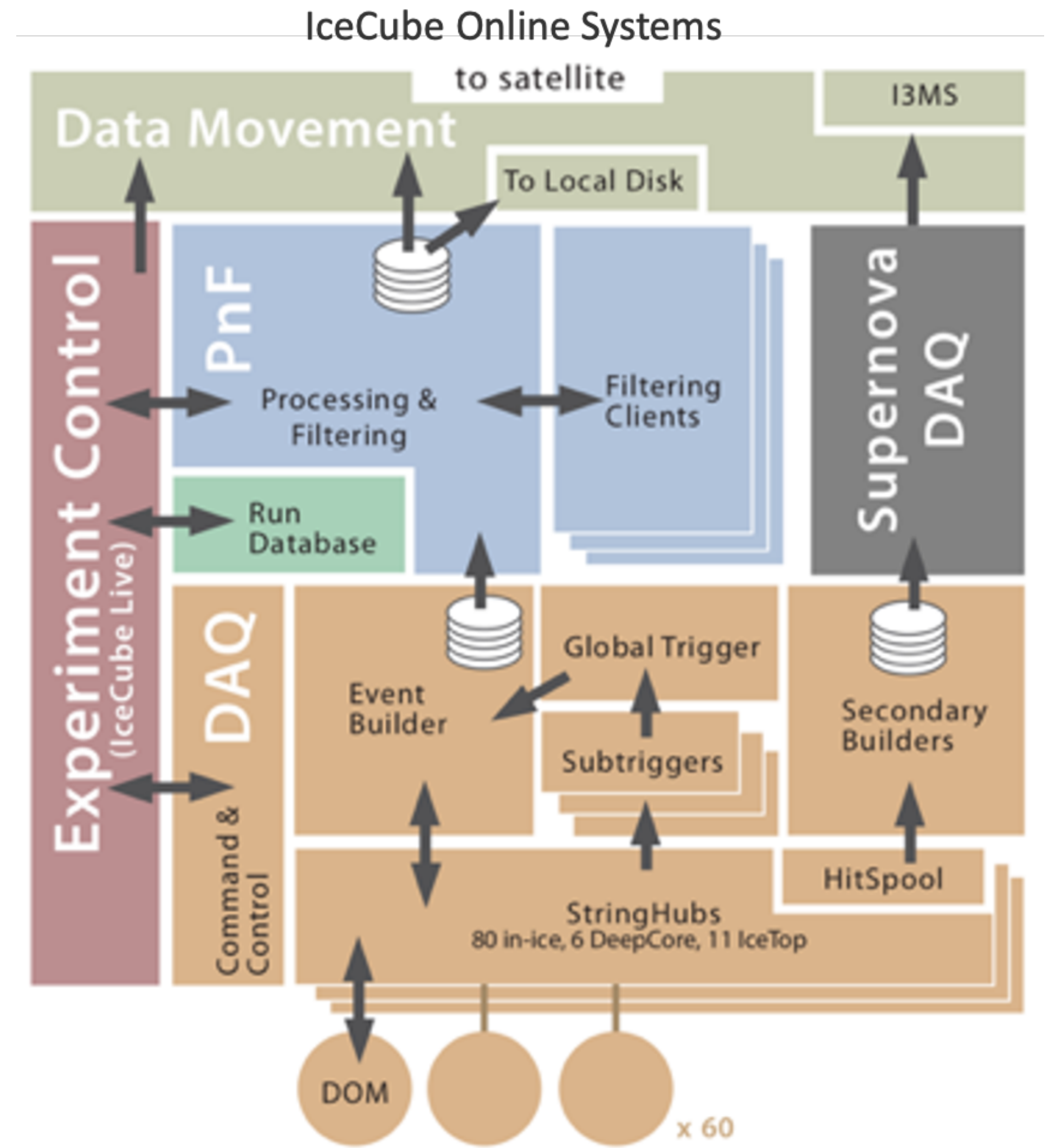
Note: Much of this hasn't changed in a LONG time...



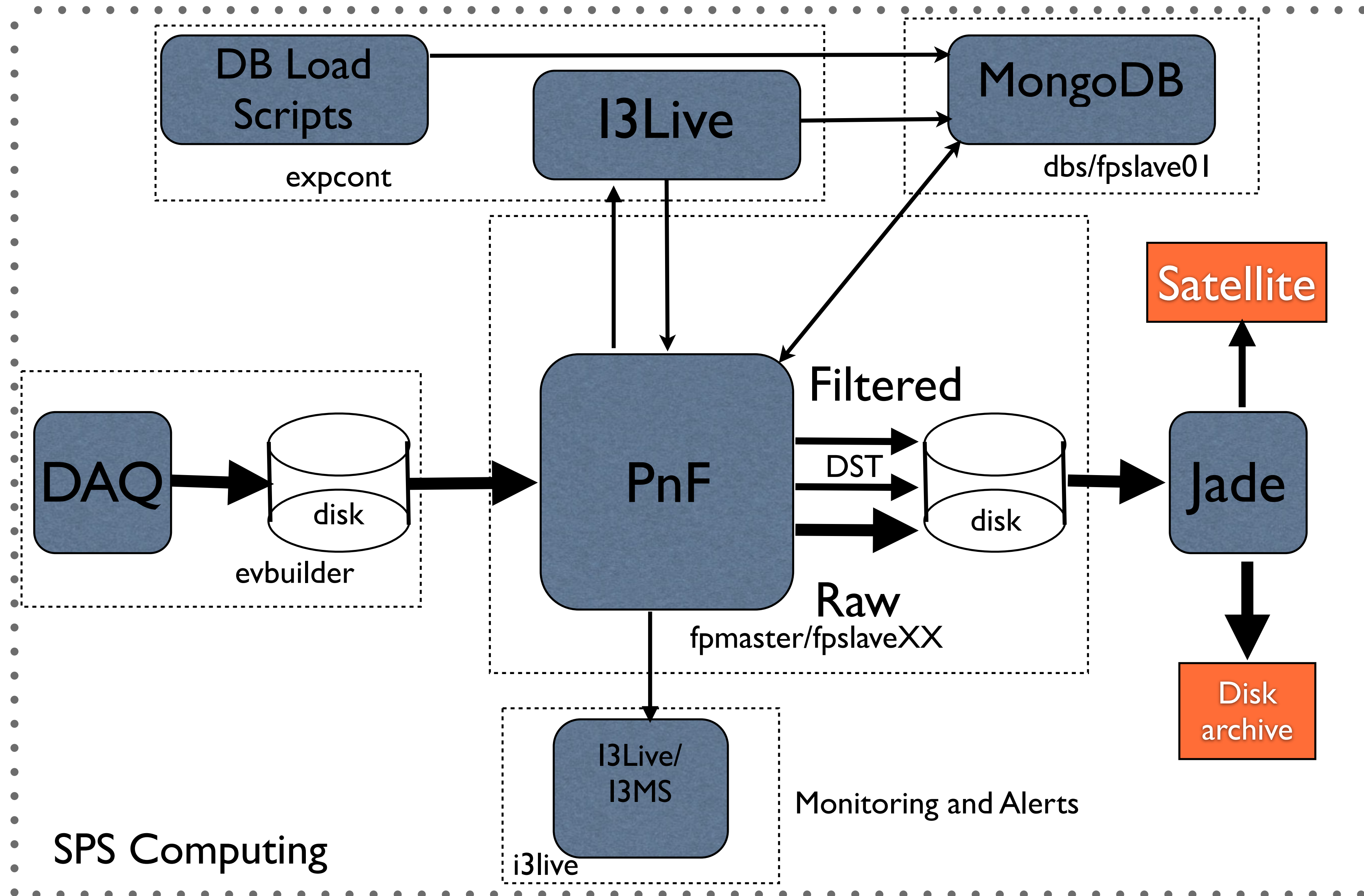
IceCube online



- Online Processing and Filtering system
 - AKA “PnF”
 - Processes ALL events from IceCube DAQ
 - Applies “LI” filters
 - All events saved to HDD arrays
 - All selected events North via TDRS
 - Realtime alerts
 - Sent to community within ~1 minute



PnF within IceCube



Data Acquisition

131518 RUNNING 1h 00m
Current run pDAQ state Duration

Tyrannena4 9,337,208 dark
pDAQ release Event count LID mode

SwitchRun 1 5409 / 5409
Transition Active DOMs

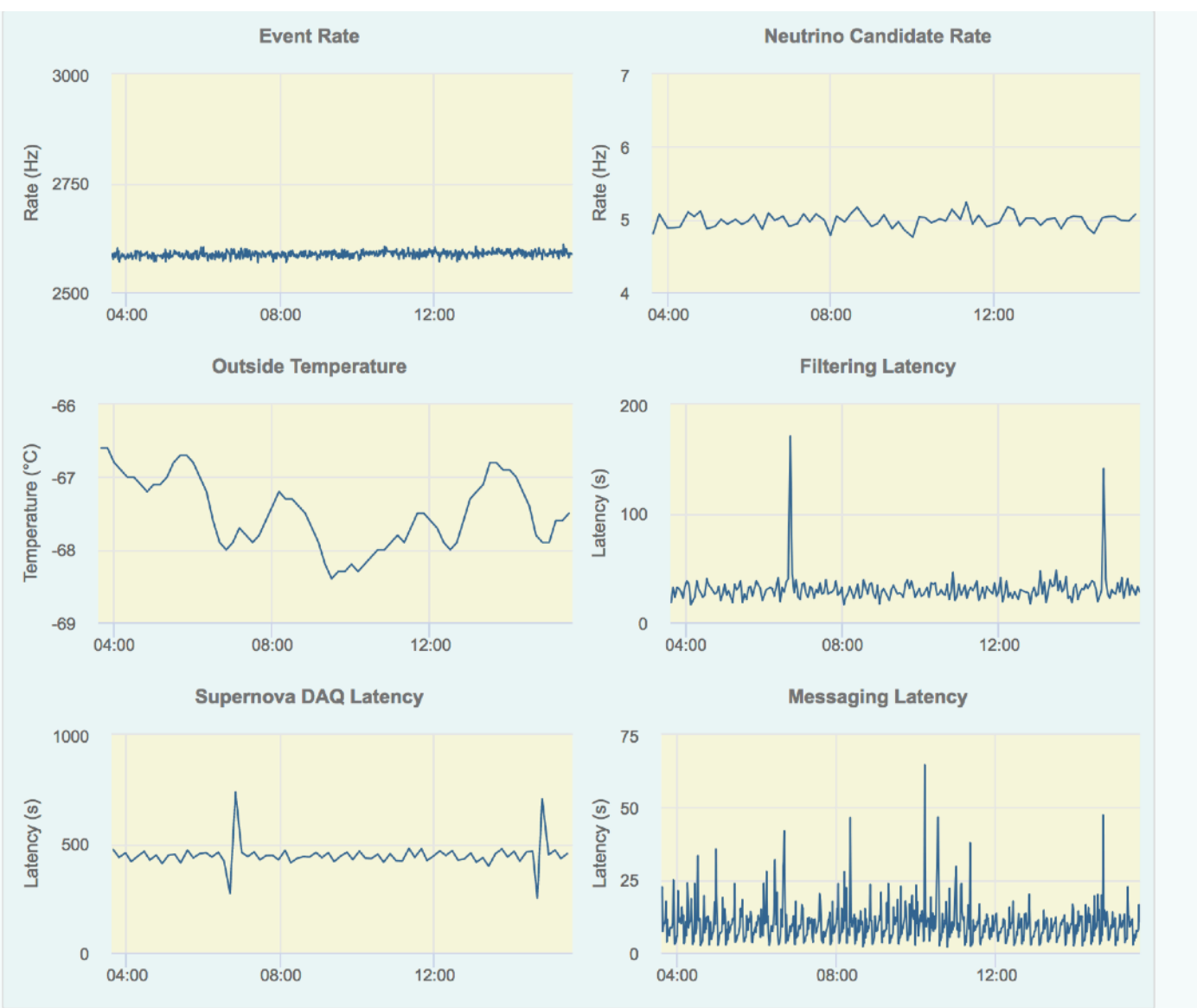
sps-IC86-2018-tweak-Grilled-Cheese-V279
Run configuration

Components

I3DAQDispatch RUNNING	PFDSTWriter RUNNING
PFFiltWriter RUNNING	PFMoniWriter1 RUNNING
PFMoniWriter2 RUNNING	PFMoniWriter3 RUNNING
PFMoniWriter4 RUNNING	PFOneWriter RUNNING
PFRawWriter RUNNING	PFSer1 RUNNING
PFSer2 RUNNING	PFSer3 RUNNING
PFSer4 RUNNING	PFSer5 RUNNING
PFSer6 RUNNING	PFSer7 RUNNING

Live Alerts

- multirunfail
- runfail
- FPMaster disk usage too high
- PnF file backlog too large
- Too much time betw...DAQ latency messages
- Detector not taking data
- Excessive run failure rate
- I3DAQDispatch is not RUNNING
- I3MS link is not UP
- Lots of LBM overflows
- Modem ttyS1 is unable to connect
- Modem ttyS4 is unable to connect
- Modem ttyS5 is unable to connect
- Modem ttyS6 is unable to connect
- PFDSTWriter is not RUNNING
- PFFiltWriter is not RUNNING
- PFRawWriter is not RUNNING
- PFSer1 is not RUNNING
- PFSer2 is not RUNNING
- PFSer3 is not RUNNING
- PFSer4 is not RUNNING
- PFSer5 is not RUNNING
- PFSer6 is not RUNNING
- PFSer7 is not RUNNING
- PFSer8 is not RUNNING
- PnF latency too high
- PnF rate too low
- PPP link is not UP
- SERIOUS SN alert triggered!
- SNDAQ is not RUNNING
- SNDAQ latency too high
- Throttling messages from noisy service
- Too many *.dat file...local/pdaq/sndaq/tmp
- Too many *sn.tar fi... /mnt/data/pdaqlocal



Recent Log Msgs

```

2018-09-18 15:10:28 Watchdog reports starved components: stringSub#201->IceTopTrigger icetopHit.RecordsReceived not changing f
2018-09-18 14:36:44 Wrote combined log for run 131517
2018-09-18 14:36:37 Writing combined log for run 131517
2018-09-18 14:36:25 Run switched SUCCESSFULLY.
2018-09-18 14:36:25 74414961 physics events collected in 28757 seconds (2587.72 Hz)
2018-09-18 14:36:21 Switching to run 131518...
2018-09-18 14:36:21 Cluster: sps-cluster
2018-09-18 14:36:21 Run configuration: sps-IC86-2018-tweak-Grilled-Cheese-V279
2018-09-18 14:36:21 Version info: Tyrannena4 14222:143282 2018-08-10 20:20:07
2018-09-18 14:36:21 pdaq: RUNNING -> switchrun {'runNumber': 131518, 'subRunNumber': 0, 'runConfig': 'sps-IC86-2018-tweak-Gril
2018-09-18 13:38:53 Watchdog reports starved components: stringSub#201->IceTopTrigger icetopHit.RecordsReceived not changing f
2018-09-18 13:17:02 Watchdog reports starved components: stringSub#201->IceTopTrigger icetopHit.RecordsReceived not changing f
2018-09-18 13:01:54 Watchdog reports starved components: stringSub#201->IceTopTrigger icetopHit.RecordsReceived not changing f

```

Status Recent Systems History Comms Help

PnF Recent History

Active analysis clients:

server 1: **51** server 2: **51** server 3: **51** server 4: **51**
server 5: **51** server 6: **51** server 7: **51** server 8: **51**

Current Filters are overlapping, needlessly complex, and a mystery box

Neutrino Sources

- GFUFilter_17
- GRECOOnlineFilter_19
- MuonFilter_13
- OnlineL2Filter_17

Cosmic Ray

- IceActTrigFilter_18
- IceTop_InFill_STA2_17
- IceTop_InFill_STA3_13
- IceTopSTA3_13
- IceTopSTA5_13
- InIceSMT_IceTopCoincidence_13
- MoonFilter_13
- ScintMinBias_16
- SDST_IceTop_InFill_STA3_13
- SDSTIceTopSTA3_13
- SDST_InIceSMT_IceTopCoincidence_13
- SunFilter_13

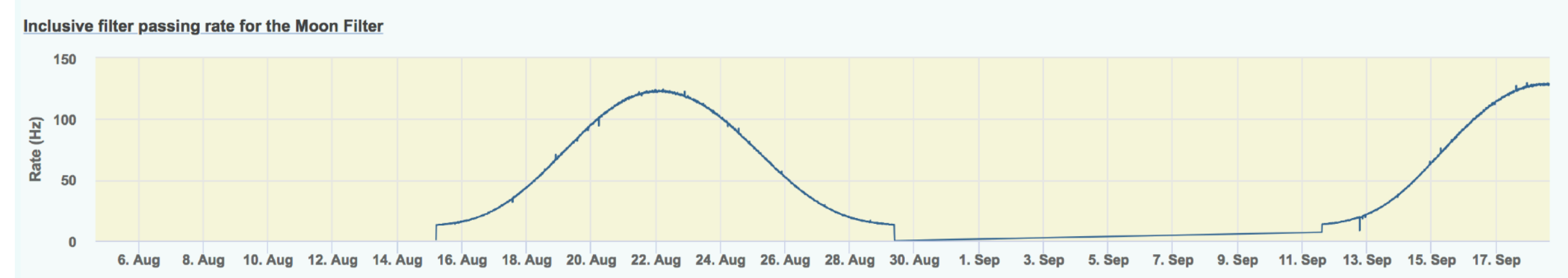
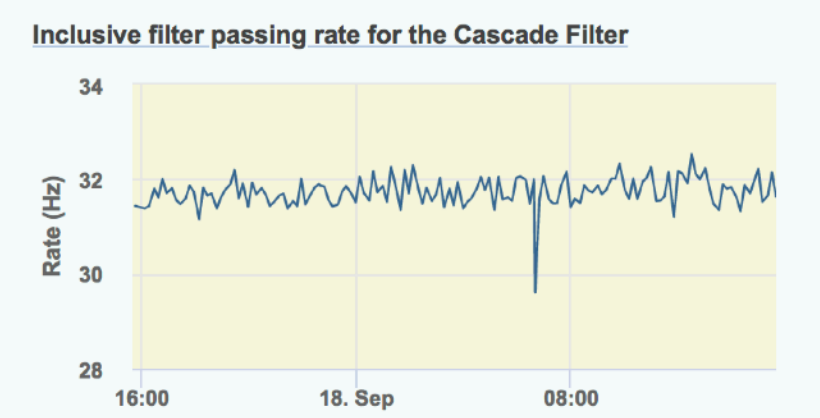
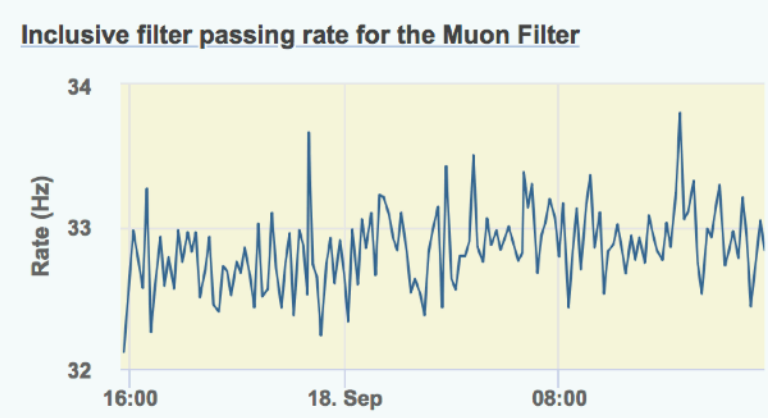
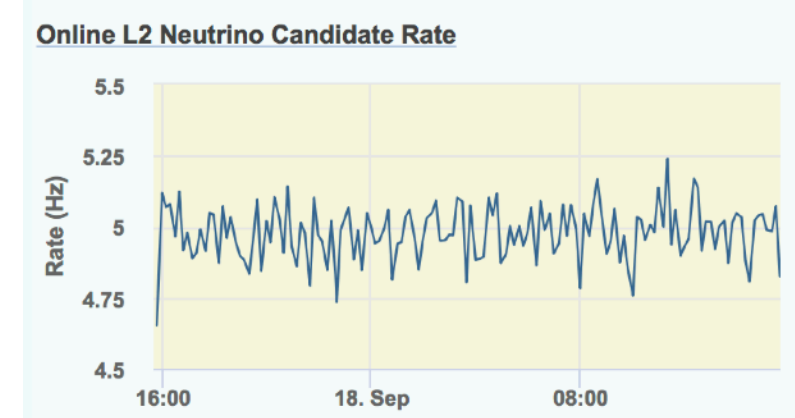
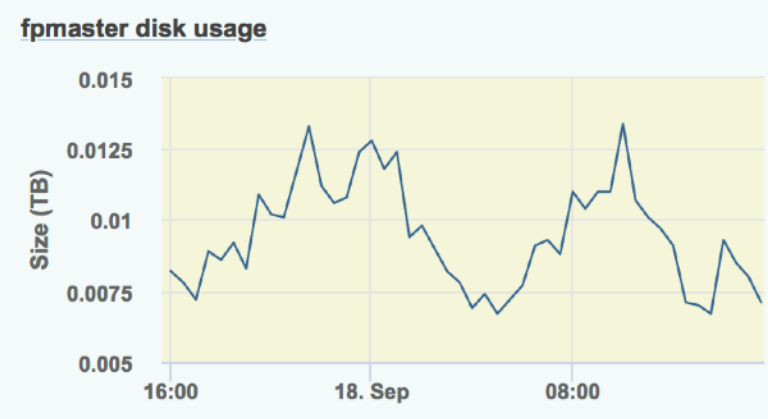
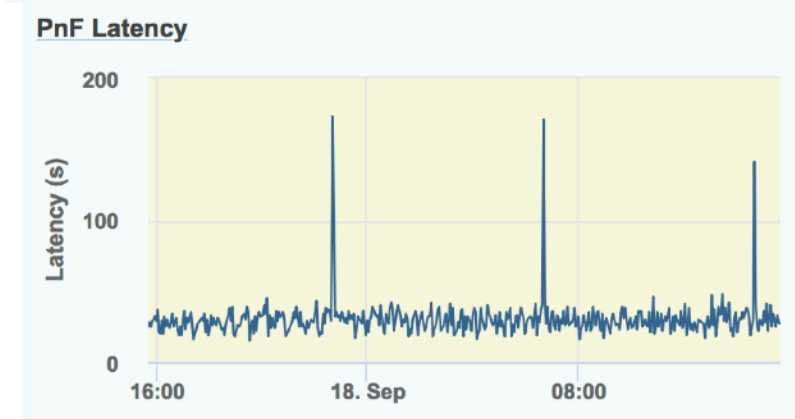
Diffuse

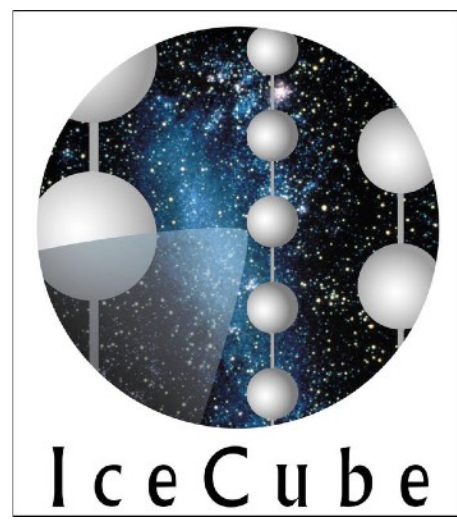
- CascadeFilter_13
- EHEAlertFilter_15
- EHEAlertFilterHB_15
- EstresAlertFilter_18
- HESEFilter_15
- HighQFilter_17
- MESEFilter_15

Beyond Standard Model

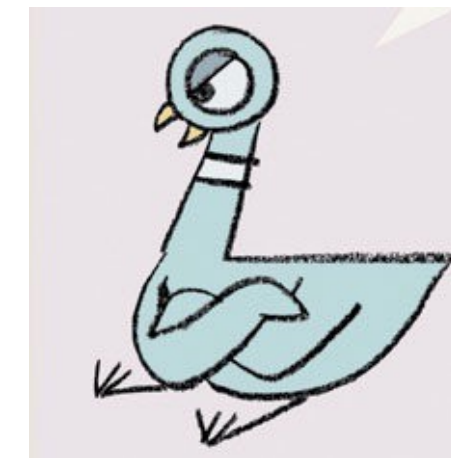
- FSSCandidate_13
- FSSFilter_13
- LowUp_13
- MonopoleFilter_16
- SlopFilter_13
- VEF_13

Filter	2020 test run [Hz]	2020 test run [GB/day]
FSSFilter	187.3	7.3
MoonFilter	100.0	4.3
SDST_InIceSMT_IceTopCoinc	57.6	3.2
HighQFilter	0.8	2.6
SLOPFilter	10.9	2.5
MuonFilter	34.5	1.8
MonopoleFilter_16	30.1	1.4
CascadeFilter	33.2	1.2
IceTopSTA5	1.3	1.1
LowUp	27.6	0.9
FilterMinBias	2.7	0.8
MESEFilter_2015	9.4	0.6
DeepCore	16.2	0.5
InIceSMT_IceTopCoinc	0.7	0.4
ICOnlineL2Filter	5.2	0.4
VEFFilter	12.0	0.4
IceTopSTA3	0.6	0.4
SDST_IceTopSTA3	5.9	0.2
FixedRateFilter	0.0	0.1
ScintMinBias_16	2.2	0.1
IceTop_InFillSTA3	0.1	0.1
IceTop_InFill_STA2_17	2.4	0.1
SDST_IceTop_InFill_STA3	1.2	0.0



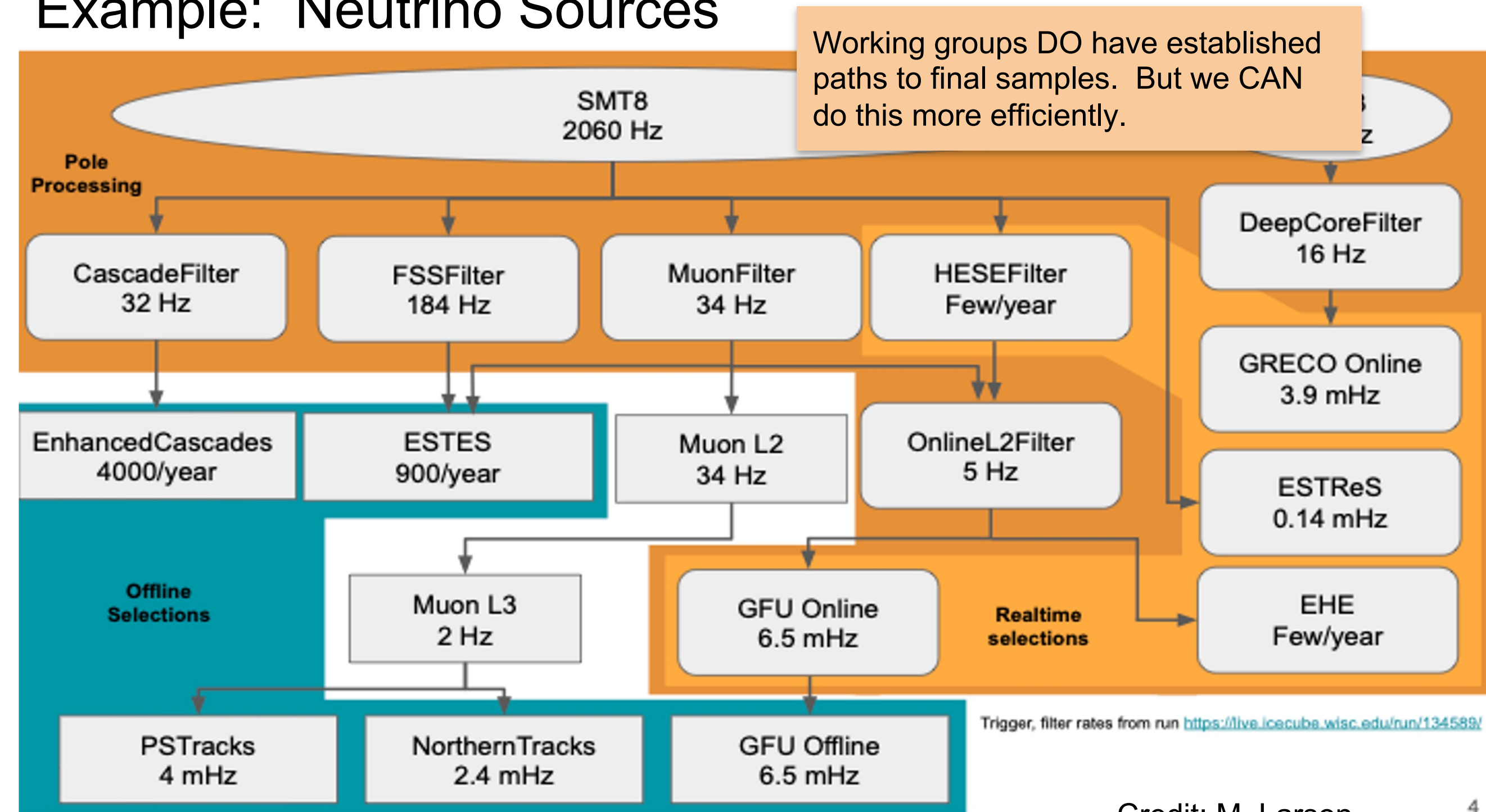


Higher Level Filters



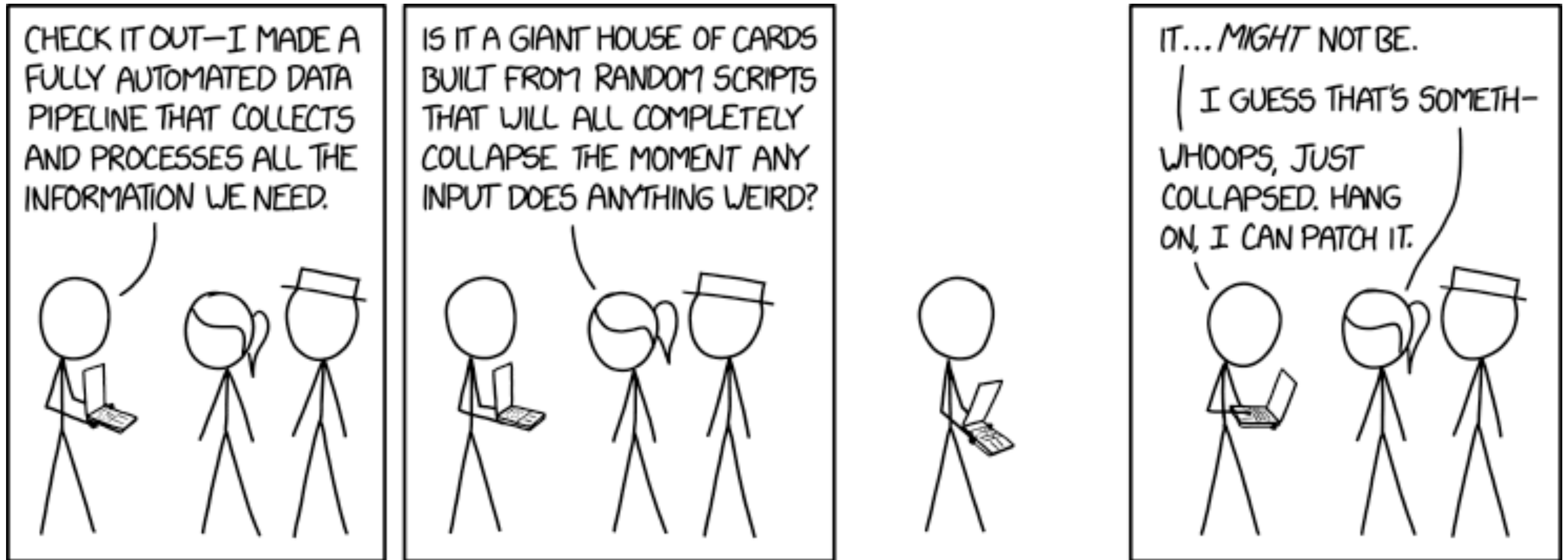
- L2 filters - Common processing
 - Expands L1 filter files from Pole
 - Adds additional reconstructions, etc
 - Removes nothing
- L3 Filters - Topical (tracks, cascades, etc) specific
 - Select ~few Hz of most interesting events for additional processing.
- WG specific process - generate “final samples”
 - Can be multiple steps (L6?) or draw from many sources

Example: Neutrino Sources



Credit: M. Larson

Moving forward - new plan



Obligatory XKCD..

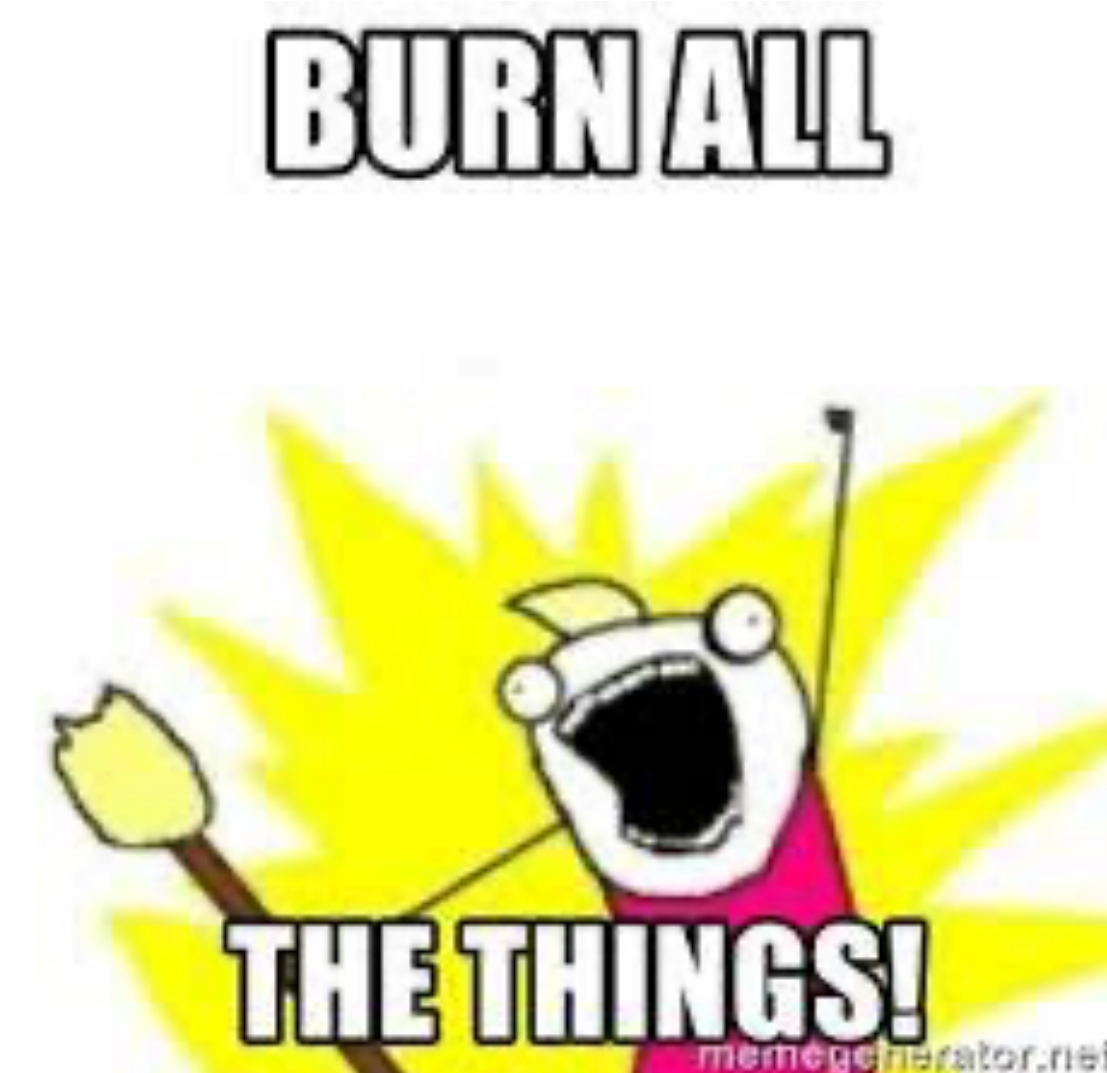
What are we talking about here...

For a couple of years now we've been talking about making large changes to our existing filtering and data reduction pipelines.

- New additions, bug fixes, etc to existing software (called Pass3 in the past)
- Reduce complexity of existing software
- Reduced resource consumption

Many of us feel it's time for a total rethink

- **Clearer** code and easier software maintenance going forward
- **Reduced** resource consumption in data processing and reduction
 - Cut harder, earlier.
- **Faster** to physics.
- Be better prepared for **Upgrade(s)**



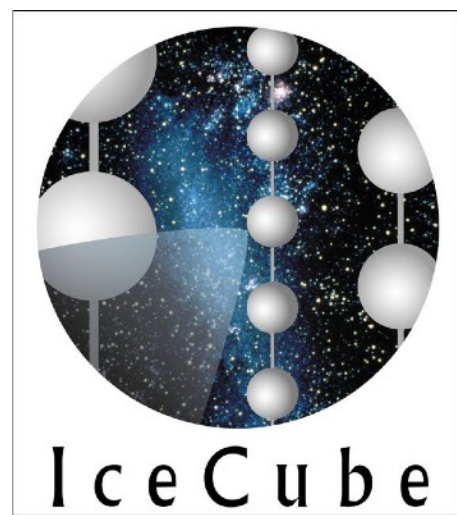
See [Naoko's talk](#) at the last collaboration or [my talk at ICC](#) meeting for more...

New filter strategy

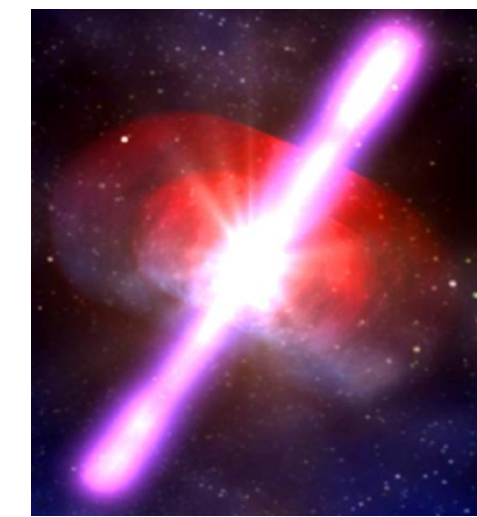


At Pole:

- Apply latest calibration and feature extraction tools (“Pass3”) to all events
- Create SuperDST record for all events (~160 GB/day) and send North via TDRS
 - Likely some room for full waveform data for selected events based on WG need
 - Currently testing: “HighQ Filter” at 1000pe as initial “Save all waveforms” flag
 - IceTop saves all IceTop launches for “large” events
 - Gives us a compact record of all (most) events in the north quickly.
- Apply realtime filter selections at pole
 - Select alert events, online PS sample (aka GFU selection) in realtime
 - Send event summaries and event data north via I3MS/Iridium for immediate use
 - Very much a work in progress
- Perform detector data monitoring
 - Report histograms to I3Live as currently done

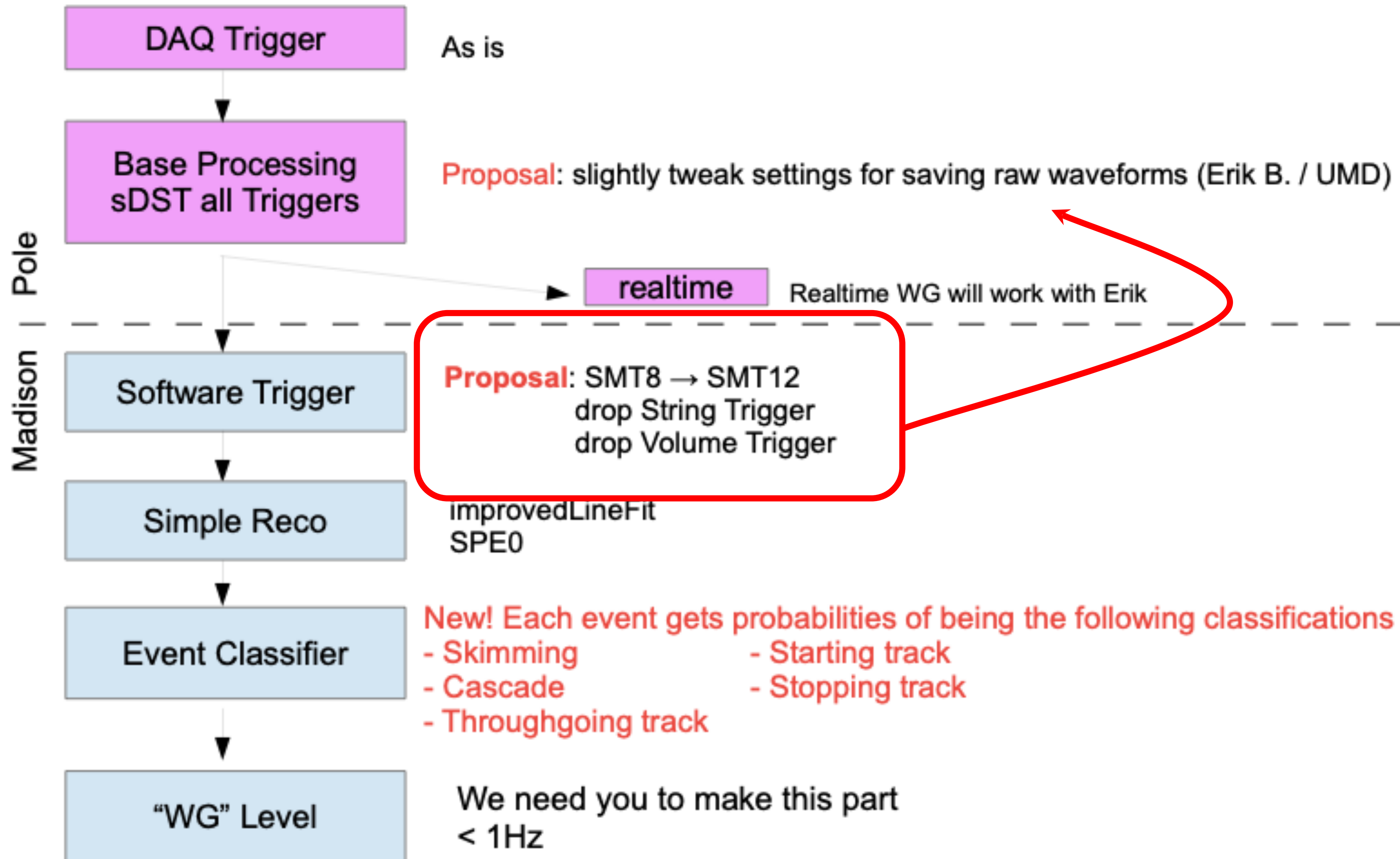


SuperDST



- SuperDST is a compact representation of the pulses extracted online after calibrating and feature extracting waveforms collected online.
- Encoding of Time/Charge of all pulses from all DOMS
- Nicely integrated into IceTray: Can read directly as I3RecoPulses...
- More information [I3Wiki:SuperDST](#)
- Seatbelts, please. We save inice waveforms (DAQ format) whenever (2013-2015 criteria):
 - Original waveform does not match a SuperDST “rebuilt” waveform
 - >1 ATWD active
 - >1000 pe in 1 usec
 - These can be used to recalibrate, re-extract pulses with update calibrations, new algorithms, etc

New Online & Offline Filtering Proposal



From Naoko's collab meeting talk

Status of Northern Filtering

- Working groups are working to update their Northern processing
 - Target ~1 Hz of output.
 - Two approaches suggested:
 - **Freshen existing filtering**
 - Clean up ancient reconstructions, variables that aren't used, data streams no one looks at, etc.
 - Most working groups taking this approach – important for comparison as well.
 - NS group has a nice modernized version of the muon filter (Bennett, Georgia Tech)
 - **Redo everything!**
 - NS group rethought what they were doing and came up with super-neat modern techniques for filtering.
 - If you have time, this is even better!
 - We need to finish this up this year

https://wiki.icecube.wisc.edu/index.php/2022_Filter_Upgrade_Plan_for_L1/L2/L3

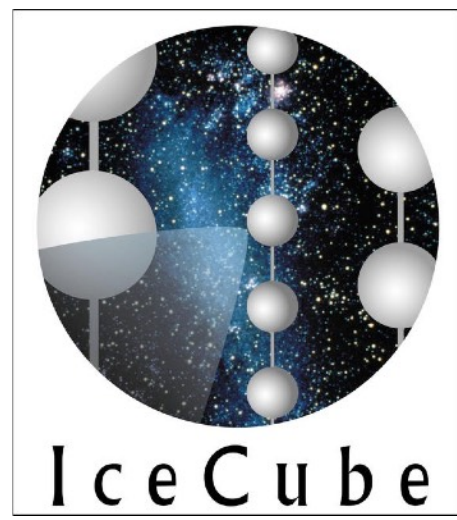
Planned timeline

None of this has been easy, since many impacts to existing analyses, etc.

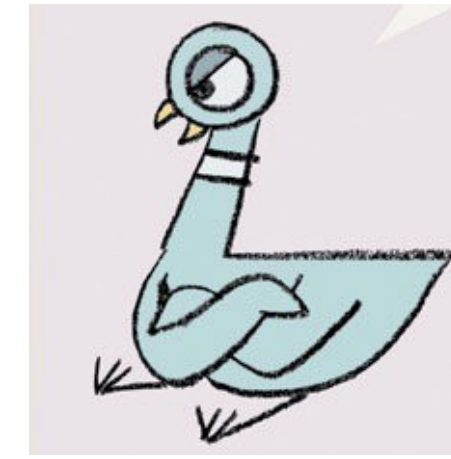
We are pushing forward

- Pole filters are being replaced this summer
- Over the Fall and Winter: WG filtering in the North will be fine tuned.
 - As new IceCuber's this is likely a place you're going to be plugging in!





Thanks!



- Questions?
- Happy to answer questions as best I can, now or later
 - @blaufuss on slack
 - blaufuss@umd.edu