

RPVLL reprocessing status

David South

Data Carousel Meeting
4 February 2020

HELMHOLTZ RESEARCH FOR
GRAND CHALLENGES



<https://www.memorabletv.com/features/classic-tv-revisited-magic-roundabout/>



RPVLL reprocessing

Some basics

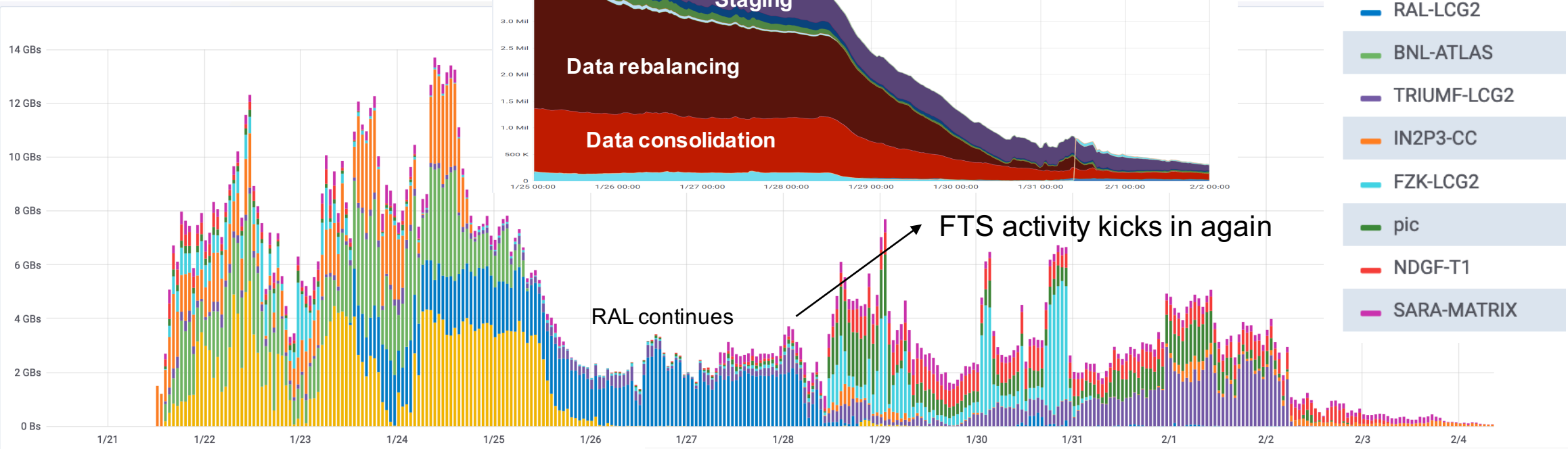
- Reprocessing of complete run 2 (2015-2018) data, 596 runs in total
- Output only DRAW_RPVLL rather than AOD, but nevertheless needs to read all RAW input
- Perform in “data carousel” mode, to avoid staging all inputs data in advance (total volume is around 18PB)

- Once DRAW_RPVLL is produced, use this to make (merged) DAOD_RPVLL
- DRAW_RPVLL format is also zipped to archive in separate step

Year	Input Size (PB)	Runs
2015	1.04	66
2016	5.28	150
2017	5.71	186
2018	6.24	194

- Only 2018 submitted so far; 2017, 2016 and finally 2015 still to come
- 2018 request: https://prodtask-dev.cern.ch/prodtask/inputlist_with_request/29088/

Input staging



- Most staging of 2018 data now completed, tail of a few hundred MB at SARA-MATRIX
- CERN-FTS situation relieved on 28.01, sites relying on this service kicking in again after this

Source	Progress plot	DDM Dash	Datasets Queuing	Datasets Active	Datasets Done (+ 90% readiness)	Files Remaining	Files Queuing	Files Done
CERN-PROD_RAW	✓	-->	0	0	29 (+0)	0	0	501292
IN2P3-CC_DATATAPE	✓	-->	0	5	16 (+0)	16543	0	253114
FZK-LCG2_DATATAPE	✓	-->	0	0	14 (+0)	0	0	215129
SARA-MATRIX_DATATAPE	✓	-->	7	14	0 (+6)	140803	114573	137220
NDGF-T1_DATATAPE	✓	-->	0	12	0 (+12)	274	0	172465
RAL-LCG2_DATATAPE	✓	-->	0	1	24 (+1)	749	0	358312
PIC_DATATAPE	✓	-->	0	6	4 (+6)	95	0	171151
TRIUMF-LCG2_DATATAPE	✓	-->	0	12	2 (+12)	37	0	250492
BNL-OSG2_DATATAPE	✓	-->	0	0	33 (+0)	0	0	395610

Reprocessing progress

2018 status

https://prodtask-dev.cern.ch/dkb/#/output_stat/?hashtag=%7Cdraw_reprocessing012020

Hashtags: |draw_reprocessing012020

total	staging	registered	assigning	submitting	ready	running	done
776	20	394	16	16	119	83	128

Tasks with at least 'ready' status('staging' and 'registered' are not included):

Step Name	DRAW_RPVLL r11782 43.13%	DRAW_RPVLL p4071 74.72%	DAOD_RPVLL r11784 61.68%	DAOD_RPVLL p4072 96.99%
Tasks	174	66	121	9
Input events	5,066,477,326	158,089,742	~40,573,439	14,012,596
Processed events	2,185,282,261	118,129,490	25,025,447	13,591,112
Output events	234,668,343	118,113,021	25,025,447	13,591,112
Running/Pending	18%/39%	14%/11%	1%/38%	3%/0%
Input bytes	4.99 PB (174 tasks)	170.06 TB (66 tasks)	249.39 TB (120 tasks)	14.92 TB (9 tasks)
Input bytes (done tasks)	1.55 PB (67 tasks)	96.13 TB (43 tasks)	15.38 TB (9 tasks)	7.04 TB (7 tasks)
Output bytes	252.55 TB (120 tasks)	128.04 TB (63 tasks)	27.05 TB (25 tasks)	14.46 TB (9 tasks)
Average HS06 per event	588	5	2042	38
Duration (finished tasks)	5.02 days	1.12 days	5.89 days	0.93 days

RAW to DRAW_RPVLL

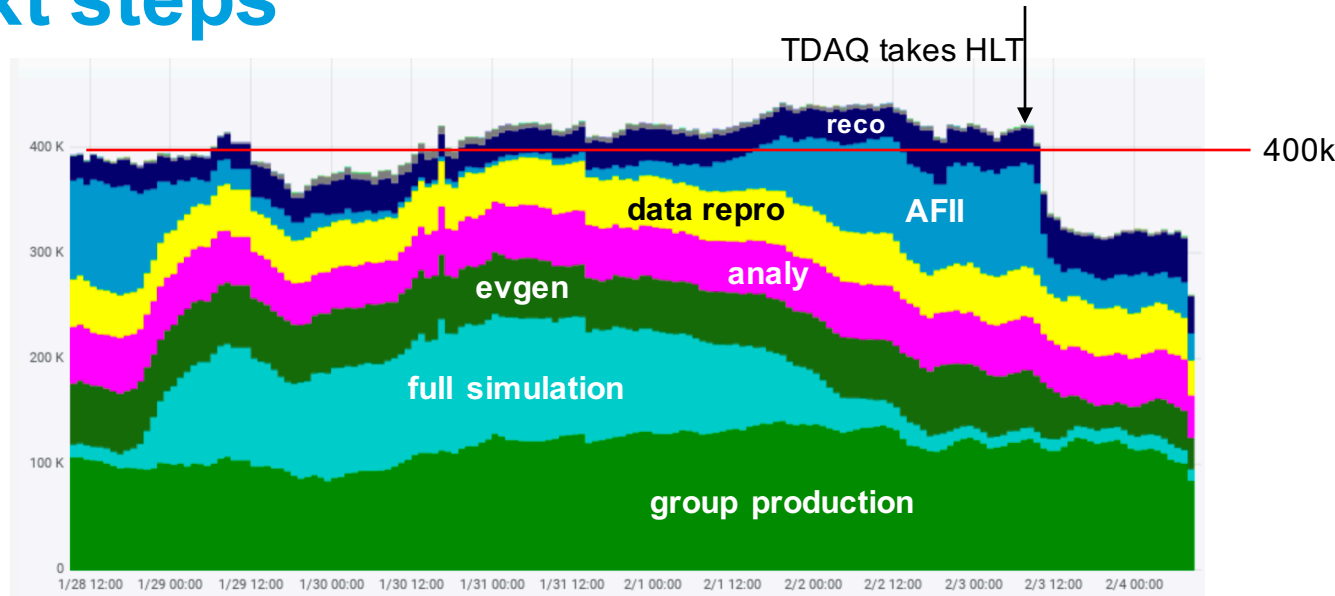
DRAW_RPVLL archive

DRAW_RPVLL to
DAOD_RPVLL

DAOD_RPVLL merge

- 2018: 194 runs, 174 have started, remainder still staging
- Raw to DRAW_RPVLL task of 67 runs has now finished
- Approaching 50% of 2018 done in around 2 weeks
- Original estimate of 4 weeks for 2018 looks about right

Next steps



By Source Tape

Tape	Rules to delete	Size
BNL-OSG2_DATATAPE	23	491.62 TB
FZK-LCG2_DATATAPE	9	274.55 TB
IN2P3-CC_DATATAPE	6	160.39 TB
PIC_DATATAPE	2	20.56 TB
RAL-LCG2_DATATAPE	14	390.15 TB
total	54	1.31 PB

By Destination Disk

Datadisk	Rules to delete	Size
AGLT2_DATADISK	3	97.14 TB
BNL-OSG2_DATADISK	10	166.02 TB
FZK-LCG2_DATADISK	5	135.87 TB
IN2P3-CC_DATADISK	5	149.85 TB
IN2P3-LAPP_DATADISK	1	10.54 TB
LRZ-LMU_DATADISK	1	19.12 TB
MWT2_DATADISK	2	53.06 TB
NET2_DATADISK	4	70.33 TB
PIC_DATADISK	2	20.56 TB
PRAGUELCG2_DATADISK	1	41.57 TB
RAL-LCG2-ECHO_DATADISK	9	252.82 TB
SWT2_CPB_DATADISK	4	105.07 TB
UKI-LT2-QMUL_DATADISK	3	88.79 TB
UKI-NORTHGRID-MAN-HEP_DATADISK	2	48.53 TB
UNI-FREIBURG_DATADISK	2	77.99 TB
total	54	1.31 PB

- Reprocessing using around 10% of resources, enough for now
- 2018 should be done in ~ 2 weeks, consider week after S+C week for next round with 2017 (similar size, runs to 2018)
- Tables show data volume of used RAW inputs, currently ~ 1.3PB
 - Default lifetime of 30 days (~20 Feb), in addition current “carousel procedure” will delete RAW rules only when whole chain is done (all 4 steps)
 - Consider changing that to when the DRAW is 100% done
 - Note: CERN-PROD inputs (from INFN tape in Aug '18) have infinite lifetime