

Weekly report

Date: 05/23/2019

Speaker: Andrew Liu

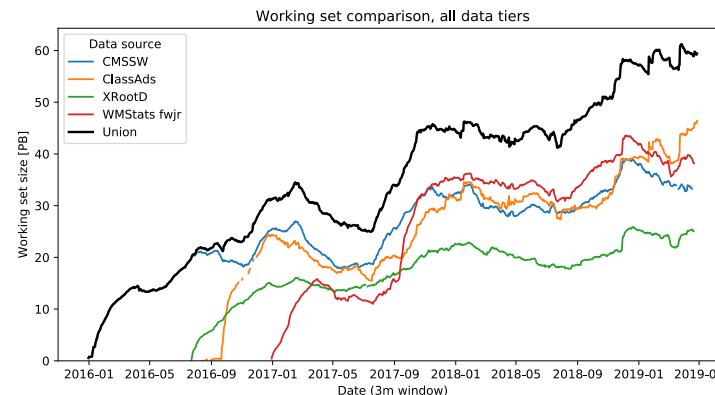
Outline

- Number of blocks per day and per job for each working set
- Current progress on tracing the discrepancy between the average number of blocks of ClassAds and that of the other four working sets
- Goal of next week

Number of blocks per day and per row for each working set

	ClassAds	CMSSW	fwjr	jmpopularity	XrootD
Number of blocks per day	1598.82	61676.81	108982.26	50116.58	110216.59
Number of blocks per row	11.76	42.19	146.48	41.92	21.77

- Do I need to filter the working sets first and then analyze the filtered working sets?



Column name of each working set

ClassAds

	day	input_campaign	d_data_tier_id	working_set
0	1.473379e+09	RunIISummer15GS	109	[12386981]
1	1.473725e+09	Run2016B-v2	1	[12896869]
2	1.474848e+09	RunIIFall15DR76	118	[12412865]

CMSSW

	day	isCrab	input_campaign	d_data_tier_id	SITE_NAME	working_set_blocks
0	1.451606e+09	None	Run2015C-v1	1	T1_DE_KIT	[14209448, 14229273, 14238958, 14212944, 14212...
1	1.451693e+09	True	RunIISpring15DR74	21	T2_DE_DESY	[14349037]
2	1.451693e+09	True	RunIISpring15MiniAODv2	31224	T2_DE_DESY	[14450122, 14465838, 14448750, 14442718, 14450...

fwjr

	day	input_campaign	d_data_tier_id	site_name	working_set_blocks
0	1483142400	RunIISummer15wmlHEGS	109	T2_ES_CIEMAT	[16971024, 16971297, 16980105, 16972135, 16971...
1	1483228800	RunIISpring15PrePremix	113	T1_RU_JINR	[16523829, 16563094, 16578077, 16555465, 16576...
2	1483228800	RunIISpring15PrePremix	113	T2_DE_DESY	[16573192, 16572705, 16580365, 16548192, 16579...

jmpopularity

	day	SubmissionTool	input_campaign	d_data_tier_id	SiteName	working_set_blocks
0	1.453075e+09	crab3		HC	122	T2_FR_GRIF_LLR [12116542, 12117213]
1	1.453075e+09	crab3		HIRun2015	9	T2_US_Nebraska [14918229]
2	1.453162e+09	crab3	Run2015D-16Dec2015	31223	T2_CH_CERN	[14830368, 14819763, 14818768, 14823581, 14908...

XrootD

	day	input_campaign	d_data_tier_id	client_domain	working_set_blocks
0	1.382400e+06	CMSSW_8	141	cern.ch	[16053141]
1	1.382400e+06	Fall13	109	physik.rwth-aachen.de	[11289134, 11289505, 11289698, 11289495, 11289...
2	1.382400e+06	HC	122	1.1.219]	[12116542, 12117213]

Current progress on tracing the discrepancy of the average number of blocks of ClassAds

- 1. The data size of each parquet file in ClassAds is around 30-40 KB and that in the other working sets is around 1 MB, which is 30-40 times larger than that in ClassAds (each working set has same number, 202, of parquet files).
- 2. In ClassAds, the column name for working set blocks is "working set", but in the other working sets it is "working set blocks". Maybe in Nick's ntuple, the definition of working set blocks column for ClassAds is different from that of the other working sets?

Goal of next week

- 1. Get access to the data used to plot this figure and then replot the figures I produced last week

