Proposal for Sharing of Data on Data Access

Diego Davila, June 2019
Outline

- The Goal
- The 3 datasets we propose
- Where is the data? and What is available?
- How are they grouped/compressed?
- How they look? and How big they are?
The Goal

To create a set of datasets that:
- can be used to understand and find patterns of data access
- can be shared among the WLCG collaboration
- is minimal in size
The 3 datasets that we propose

1. **input_data**: This dataset describes the data accessed by the jobs
2. **analysis_jobs**: Describes the ‘analysis’ jobs that read data
3. **production_jobs**: Describes the ‘production’ jobs that read data
The Data Bookkeeping Service (DBS) provides a catalog of event metadata for Monte Carlo and recorded data of CMS.

- Comprises all necessary information for tracking datasets, their processing history and associations between runs, files and datasets.
- All kind of data-processing as well as physics analysis done by the users are heavily relying on the information stored in DBS.
input_data - DBS structure

- **Datatier.** 63 different
  AOD, MINIAOD, GEN-SIM
- **Dataset.** +900K different
  /EGamma/Run2018A-17Sep2018-v2/MINIAOD
- **Block.** +9.5M different
  /EGamma/Run2018A-17Sep2018-v2/MINIAOD#c460460b-a4ac-454a-ab42-723e6c418826
- **File.** +135M
  /store/data/Run2018A/EGamma/MINIAOD/17Sep2018-v2/100000/FBEED00E-DB6B-E948-A774-936B3074776A.root
input_data - DBS tables

Datatiers
- data_tier_id
- data_tier_name
- data_tier_creation_date
- data_tier_create_by

Datasets
- d_dataset_id
- d_dataset
- d_is_dataset_valid
- d_primary_ds_id
- d_processed_ds_id
- d_data_tier_id
- d_dataset_access_type_id
- d_acquisition_era_id
- d_processing_era_id
- d_physics_group_id
- d_xtcrosssection
- d_prep_id
- d_creation_date
- d_create_by
- d_last_modification_date
- d_last_modified_by

Blocks
- b_block_id
- b_block_name
- b_dataset_id
- b_open_for_writing
- b_origin_site_name
- b_block_size
- b_file_count
- b_creation_date
- b_create_by
- b_last_modification_date
- b_last_modified_by

Files
- f_file_id
- f_logical_file_name
- f_is_file_valid
- f_dataset_id
- f_block_id
- f_file_type_id
- f_check_sum
- f_event_count
- f_file_size
- f_branch_hash_id
- f_adler32
- f_md5
- f_auto_cross_section
- f_creation_date
- f_create_by
- f_last_modification_date
- f_last_modified_by
input_data - The fields that are kept

- **b_block_id**. The unique identifier of a block
- **b_block_size**. Total size of the block in Bytes
- **ds_logical_name**. First part of the dataset name
- **ds_campaign**. Campaign name and second part of the dataset name
- **ds_campaign_sufix**. Third part of the dataset name
- **ds_datatier**. Name of the datatier, the fourth part of the dataset name
- **num_events(*)**. Sumatory of the number of events on all the files belonging to the block

(*) Grouped value
## input_data - The final dataset (159 MB)

<table>
<thead>
<tr>
<th>b_block_id</th>
<th>b_block_size</th>
<th>ds_logical_name</th>
<th>ds_campaign</th>
<th>ds_datatier</th>
<th>num_events</th>
</tr>
</thead>
<tbody>
<tr>
<td>16986783</td>
<td>8.342055e+09</td>
<td>DM_ScalarWH_Mphi-1000_Mchi-450_gSM-1p0_gDM-1p0...</td>
<td>RunII Summer 16 DR80 Premix</td>
<td>AODSIM</td>
<td>27215</td>
</tr>
<tr>
<td>19179262</td>
<td>2.158620e+07</td>
<td>JetHT</td>
<td>Run 2018 D</td>
<td>USER</td>
<td>96183</td>
</tr>
<tr>
<td>20551050</td>
<td>2.666179e+08</td>
<td>HeavyNeutrino_trilepton_M-600_V-0p01_tau_NLO_a...</td>
<td>RunII Summer 16 DR80 Premix</td>
<td>AODSIM</td>
<td>800</td>
</tr>
<tr>
<td>17630677</td>
<td>2.129800e+09</td>
<td>RelValTenMuExtendedE_200_5000</td>
<td>CMSSW</td>
<td>AODSIM</td>
<td>9000</td>
</tr>
<tr>
<td>18271220</td>
<td>3.136090e+08</td>
<td>ZeroBias3</td>
<td>Run 2017 H</td>
<td>USER</td>
<td>5946445</td>
</tr>
<tr>
<td>12108346</td>
<td>2.369607e+09</td>
<td>WminusToMuNu_M-50To250_ew-BMNNP_7TeV-powheg-py...</td>
<td>Summer 11 Leg</td>
<td>GEN-SIM</td>
<td>4400</td>
</tr>
<tr>
<td>14953094</td>
<td>2.226808e+08</td>
<td>QCD_HT1500to2000_GenJets5_TuneCUETP8M1_13TeV-powheg-py...</td>
<td>RunII Summer 15 GS</td>
<td>GEN-SIM</td>
<td>153</td>
</tr>
<tr>
<td>15884928</td>
<td>7.041365e+10</td>
<td>DisplacedJet</td>
<td>Run 2016 B</td>
<td>AOD</td>
<td>326990</td>
</tr>
<tr>
<td>16053947</td>
<td>1.260662e+10</td>
<td>QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV-powheg-py...</td>
<td>RunII Summer 15 GS</td>
<td>GEN-SIM</td>
<td>11223</td>
</tr>
<tr>
<td>18800140</td>
<td>2.786129e+09</td>
<td>HcalNZS</td>
<td>Run 2018 A</td>
<td>MINIAOD</td>
<td>84621</td>
</tr>
</tbody>
</table>
analysis_jobs - The source

- These job records are collected by a monitoring system that queries the HTCondor infrastructure of the CMS Global and Tier0 pools.
- The records are sent to an Elasticsearch instance and later on are backed up into HDFS where we read them.
- No data structure, just a set of key-value pairs (many of them!)

Notes:
- Jobs that do no read any data are not considered.
- Only jobs going through the Global and CERN pools are recorded.
- Only jobs submitted by the CRAB system will be taken into account.
**analysis_jobs** - The fields that are kept

- **day.** The day where the job was completed
- **b_block_id.** The id of the block read by the job (the link to ‘input_data’)
- **OverflowType.** The type of read done by the job:
  - FrontedOverflow - Remote from the same region
  - IgnoreLocality - remote from anywhere
  - None - Onsite
- **site_name.** The site where the job ran
- **exitCode.** One of the following: ‘Success’ or type of error: 'Environment'
  - 'Executable', 'Stageout', 'Publication', 'JobWrapper', 'FileOpen', 'FileRead', 'OutOfBounds', 'Other'
**analysis_jobs** - The fields that are kept (grouped values)

The following fields have aggregated values of all the jobs that share the same values of the previous fields: day, b_block_id, OverflowType, site_name and exitCode

- **num_jobs.** Number grouped jobs
- **sum_CpuTimeHr.** Sumatory of the CPU time per hour of all grouped jobs
- **sum_CoreHr.** Sumatory of the (WallTime * Cores) of all grouped jobs
  - a job running for 2 hrs using 4 cores will show 8 CoreHr
- **num_users.** Number of different users in the group
### analysis_jobs - The final dataset (94 MB)

<table>
<thead>
<tr>
<th>day</th>
<th>b_block_id</th>
<th>OverflowType</th>
<th>site_name</th>
<th>exitCode</th>
<th>num_jobs</th>
<th>sum_CpuTimeHr</th>
<th>sum_CoreHr</th>
<th>num_users</th>
</tr>
</thead>
<tbody>
<tr>
<td>2713</td>
<td>1.555459e+09</td>
<td>19082936</td>
<td>FrontendOverflow</td>
<td>T2_US_Vanderbilt</td>
<td>Success</td>
<td>2</td>
<td>32.478056</td>
<td>38.457500</td>
</tr>
<tr>
<td>7064</td>
<td>1.555373e+09</td>
<td>18669473</td>
<td>IgnoreLocality</td>
<td>T2_US_Caltech</td>
<td>Success</td>
<td>30</td>
<td>9.781944</td>
<td>16.243056</td>
</tr>
<tr>
<td>3508</td>
<td>1.555114e+09</td>
<td>19631682</td>
<td>None</td>
<td>T1_ES_PIC</td>
<td>Success</td>
<td>15</td>
<td>9.066944</td>
<td>9.775278</td>
</tr>
<tr>
<td>6710</td>
<td>1.555027e+09</td>
<td>19142549</td>
<td>IgnoreLocality</td>
<td>T2_CH_CERN</td>
<td>Success</td>
<td>13</td>
<td>92.464444</td>
<td>96.866944</td>
</tr>
<tr>
<td>8662</td>
<td>1.554595e+09</td>
<td>20165222</td>
<td>None</td>
<td>T2_US_Caltech</td>
<td>Other</td>
<td>2</td>
<td>0.000000</td>
<td>0.095833</td>
</tr>
<tr>
<td>9824</td>
<td>1.554509e+09</td>
<td>19085634</td>
<td>None</td>
<td>T2_US_Florida</td>
<td>Success</td>
<td>65</td>
<td>178.993889</td>
<td>241.549167</td>
</tr>
<tr>
<td>4279</td>
<td>1.556323e+09</td>
<td>19328189</td>
<td>IgnoreLocality</td>
<td>T2_UK_SGrid_Bristol</td>
<td>Success</td>
<td>12</td>
<td>19.546111</td>
<td>48.425278</td>
</tr>
<tr>
<td>9733</td>
<td>1.555373e+09</td>
<td>19994702</td>
<td>None</td>
<td>T2_US_Florida</td>
<td>Success</td>
<td>1</td>
<td>0.714722</td>
<td>0.768889</td>
</tr>
<tr>
<td>8832</td>
<td>1.554077e+09</td>
<td>17318210</td>
<td>FrontendOverflow</td>
<td>T2_US_Vanderbilt</td>
<td>Success</td>
<td>6</td>
<td>9.235833</td>
<td>10.506944</td>
</tr>
<tr>
<td>9421</td>
<td>1.554336e+09</td>
<td>18273597</td>
<td>IgnoreLocality</td>
<td>T2_UK_London_Brunel</td>
<td>Success</td>
<td>6</td>
<td>27.440833</td>
<td>58.592778</td>
</tr>
</tbody>
</table>
production_jobs

In principle this should be very similar to ‘analysis_jobs’ but given the more complicated data processes present in production jobs (i.e. different types of processing jobs) we will have to deal with that at some later time.
Questions?
References

[1] Data Bookkeeping Service 3 - Providing event metadata in CMS
   https://cds.cern.ch/record/1623287

   https://github.com/dmwm/cms-htcondor-es