# Manage the uploaded BL DCM logs as LR returned with the Raw DCM logs for MTA input

Note: simple but large files to aggregate and merge

(Initial set that we have DCM but not uploaded data was matched as best as possible. Since 18Q2, we can use day by day)

Steps (For impression, clicks & activities, run it respective and select the priority. I used impression only previously):

1. Aggregate all LR returned on 107 server:

Extract Big Lots account id from the across accounts files and append together by day

(The raw logs located in '/mnt/drv5/dcm\_logs/impressions/', etc.)

(/home/jian/analysis/Big\_Lots/Pred\_Dataset\_all\_BL\_DCMs\_and\_LR/BL\_activity\_logs/)

1. Aggregate all DCM raw logs on 107 server

(recent LR returned files with IDL: /home/jian/lr\_dcm\_biglots/BL\_LR\_download/clicks)

1. Merger together with the combined keys to identify unique line with columns:

['Event Time', 'Advertiser ID', 'Campaign ID', 'Ad ID', 'Rendering ID', 'Creative Version', 'Site ID (DCM)', 'Placement ID', 'Country Code', 'State/Region', 'Browser/Platform ID', 'Browser/Platform Version', 'Operating System ID', 'Designated Market Area (DMA) ID', 'City ID', 'ZIP/Postal Code', 'U Value', 'Event Type', 'Event Sub-Type', 'Impression ID']

(/mnt/clients/juba/hqjubaapp02/jliang/Projects/Big\_Lots/Analysis/2019\_Q4/Predictive\_Model\_Building/mapping\_unuploaded\_files\_up\_to\_20191231\_Impr.ipynb)

1. Run by day and keep the 2 columns after merge:

df\_mapping=df\_mapping[['Customer\_Link','User ID']]

write in date for reference in case 1 to many, many to 1 and many to many scenarios