

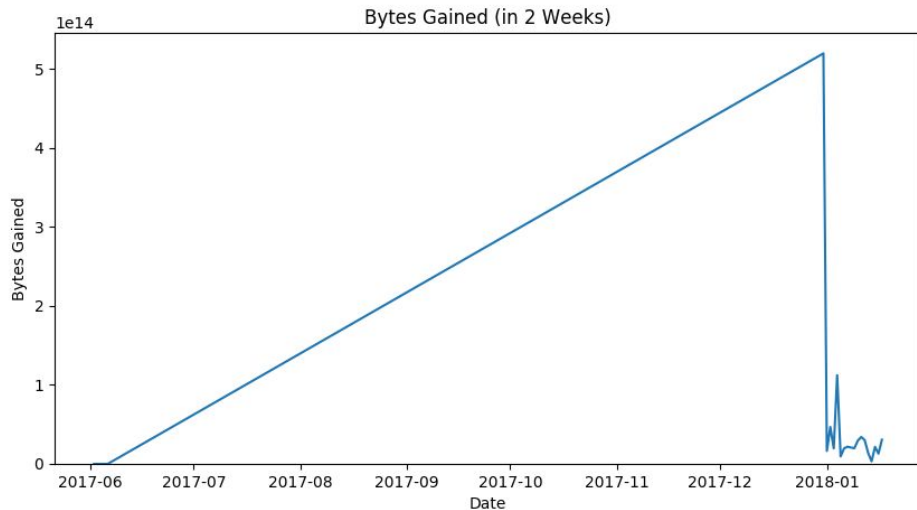
5-16-19 Weekly Report

Previous Goals

- Investigate spike (check to see if it is a bug) (Investigating blocks accessed around January 2018 individually)
- Do a characterization of the new working set
- Change to Diego's newer working set
- Ask Frank whether or not we need to filter by Crab_Jobs for the new dataset

Previous Goals

- Investigate spike (check to see if it is a bug) (Investigating blocks accessed around January 2018 individually)
- ~~Do a characterization of the new working set~~
- ~~Change to Diego's newer working set ?~~
- Ask Frank whether or not we need to filter by Crab_Jobs for the new dataset



New Dataset (5-9-16)

This is the bytes gained (within a two week threshold) per unit time for the `classAds_201801_submit.parquet` dataset.

This is for one month of Diego's newer classads working set.

Preliminary Characterization of the 2nd ClassAds Working Set

```
Number of days in working_set_day:
948
Number of days covered:
988
Dates that are covered:
2016-08-14 to 2019-04-28
Number of unique input_campaigns:
814
Total number of blocks in this classads database:
8935351
Total number of unique blocks accessed in the working set:
80922
Total number of bytes across all blocks:
180.63418737961337 PB
```

Note: There was no “njobs” column in the DataFrame so there was no way to tell how many jobs corresponded to each row.

Extensions

- Merge the months of Diego's newer working set
- Do a characterization of Diego's newer working set
- Drop everything that is not crab_job
- If in the newer working set, the spike still exists, investigate spike (check to see if it is a bug) (Investigating blocks accessed around January 2018 individually and compare to January 2019)