**Data Collection:**

1. Data has been collected from below website.
2. We have data starting from year 2013 to year 2017.
3. For each year, we have files for every day. So, that is 365 files for a year.
4. In each file, we have n number of entries, where n is the number of hard drives used on that day.

**Data Processing:**

1. Using below command, merged all the files quarter wise.

copy \*.csv **filename**.csv

1. After merging all the files, data transformation was done.

**Data Transformation:**

1. First transformation is related to the size conversion. Converted Bytes to Gigabytes.
2. Second transformation is done on the Brands of the hard drives. Replaces model names with their brand names.