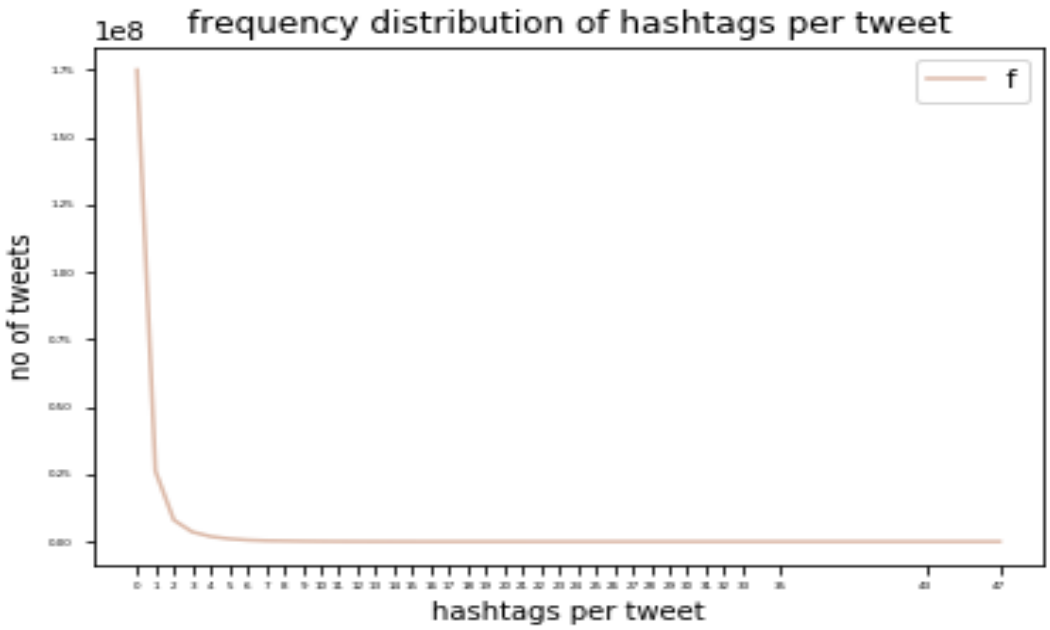
FreqTag

Logs: http://turing.cds.iisc.ac.in:8088/proxy/application_1547011148574_0180/

simple-json1.1.1 is used for parsing json.

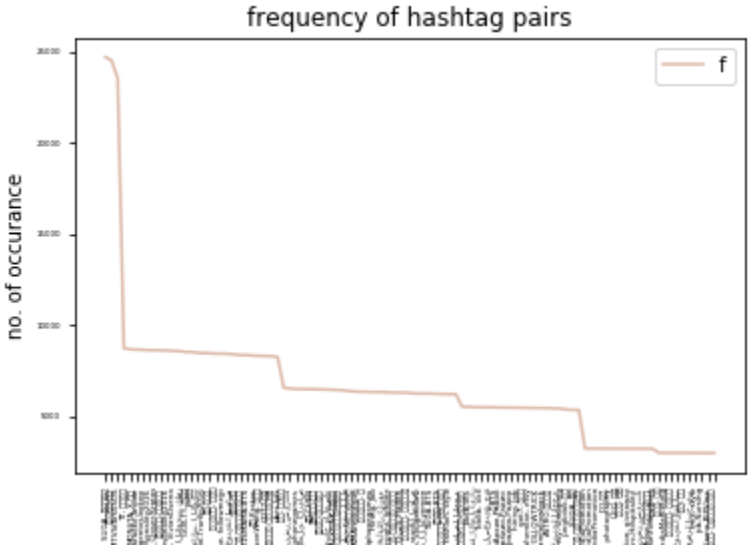
1. Filter – to filter out all the “deleted” tweets and filter out unparsable jsons.
2. map – written the no of hashtags in the tweet against each line.
3. maptopair - mapped 1 against each number obtained in previous rdd.
4. reducebykey - to obtain the frequency of each count.
5. sortbykey – sorted according to no od hashtags in a tweety.



No of executors used are 4 with 8 gb memory for each executor and 512 mb for head node with 4 executor cores.

TopCoOccurrence

Logs http://turing.cds.iisc.ac.in:8088/proxy/application_1547011148574_0182/



Operations used:

1. Filter : to filter out all the “deleted” tweets and filter out unparsable jsons.
2. Flatmap : to map all pairs of unique hashtags from a tweet.
3. Maptopair : to map 1 against all the hashtag pairs.
4. Reducebykey : to find frequency of each pair.
5. Sortbykey : to sort the hashtag pairs according to frequency.

Here are the top 100 cooccurring hashtags and their frequencies:

SUGA, 방탄소년단 24704

방탄소년단,슈가 24519

AP405SOS,MTVStars5SOS 23529

TT, 트와이스 8748

MONSTA_X,원호 8699

SingKaraoke,Smule 8674

ngentot,bokep 8664

ShaktiAstitvaKeEhsaasKi,TVPersonality2016 8644

The_Closer,Kratos 8641

오늘의방탄,방탄소년단 8629

VivianDsena,TVPersonality2016 8628

ShaktiAstitvaKeEhsaasKi, VivianDsena 8609

8592 الكويت, قطر

GUILTY, 파이터 8547

JK,정국 8545

8496 الامارات, الكويت

TrumpTrain,Trump2016 8486

호시,세븐틴 8470

방탄소년단, 제이홉 8453

followme, followmejp 8451

8441 السعودية,الرياض

제이비,재범 8406

네임드사다리,사설토토사이트추천 8376

2016MAMA,BTS 8372

빅스,Kratos 8346

JacksonWang, 갓세븐 8334

GOT7,영재 8328

사설토토추천사이트,토토사이트추천 8314

uber,lyft 8267

쯔위, 트와이스 6586

6549 صورة,نشر_سيرته

iphone,iphonegames 6516

6513 الاهلي, الرياض

종대,CHEN 6511

몬스타엑스,기현 6508

피땀눈물,방탄소년단 6496

hadith,6487 أذكار_الصباح_والمساء

BLACKPINK,블랙핑크 6471

사설토토추천사이트,네임드사다리놀이터추천 6458

6428 ايران, حزب_الله

토토,사설토토추천사이트 6391

2016MAMA, MAMAredcarpet 6366

방탄소년단, 진 6362

gameinsight,iphonegames 6351

Hiring, Job 6350

hadith,6349 أذكار_النوم

gameinsight, iphone 6328

메이저놀이터,사설토토사이트추천 6320

أذكار_الصباح_والمساء,Hadith 6315

강남야구장, 선릉풀싸롱 6311

saudi,6281 الرياض

follow, sougofollow 6269

حزب_الله,yemen 6267

SUGA,BTS 6266

君に届く,BTS 6249

부산풀싸롱, 연산동풀싸롱 6239

fashion, style 6237

6227 الرياض,الراجحي

사설토토추천사이트,스포츠토토 5549

boob,tits 5539

5520 إعادة_الأمل, ايران

boob, xxx 5516

5512 السعودية, جدة

진,BTS 5511

РадиоАрхив,Радио 5498

Архив_радио,Радио 5498

Онлайн_радио,Радио 5498

hiring, job 5490

got7, 갓세븐 5485

Bohemian, etsy 5481

MGWV,FOLLOWTRICK 5478

followback,teamfollowback 5475

ASTRO,아스트로 5467

5457 العربي,الراجحي

etsy, shopping 5443

jungkook,정국 5395

SUGA, 윤기 5375

네임드사다리,사설토토추천사이트 5371

DigitalMarketing, SocialMediaPromotion 3255

SocialMediaMarketing, SocialMediaPromotion 3255

OnlinePromotion, SocialMediaPromotion 3255

DO,디오 3254

photo, photography 3248

LEO, 빅스 3247

갓세븐, 뱀뱀 3246

competition, giveaway 3246

CareerArc,Hospitality 3243

3243 الجمعة,نشر_سيرته

Melon,멜론뮤직어워드 3242

RETWEET, TeamFollowBack 3240

아이엠, 창균 3019

SocialMedia, SocialMediaMarketing 3018

강남러시아, 강남백마 3016

3016 السعودية,الإمارات

박뿡, 지민 3014

3014 مصر,البحرين

free, style 3012

job,Nursing 3009

VineHallOfFame,RIPVine 3008

相互フォローの輪, 相互フォロー募集 3007

InterGraph

Logs : http://turing.cds.iisc.ac.in:8088/proxy/application_1547011148574_0355/

Operations used :

No of vertices :39967256

No of edges : 81555704

Operations used :

1. Filter : to filter out all the unparsable tweets and deleted tweets.
2. Map : to extract vertex information.
3. Groupbykey : to group all temporal properties of the vertices.
4. Maptopair : to extract all edges.
5. Groupbykey : to group all temporal properties of the edges.

HDFS:

Avg. block size = 114 MB

Default replication factor = 2

data-nodes =23

Head –node = 1

Total nodes = 24

Cluster Setup:

No. of Nodes = 24

RAM = 47GB

SWAP = 59GB

Data Set Description:

No. of Files = 3984 (1094 empty files are empty beginning with “tweets_1- “)

1 file 100000 tweets.

1 file size 300MB.

Spark Environment Setup:

Executor Memory = 8GB

No. of executors = 4

Driver memory =512MB