1. **Stack yang digunakan dan alur implementasi MapReduce job**

   Stack yang digunakan adalah Python, karena kemampuan Python melakukan pengolahan data dengan struktur yang tidak rumit. Sehingga dapat menghemat waktu ketika memroses input berupa JSON dengan jumlah yang cukup banyak (4074 files). Struktur file JSON yang berbeda sesuai dengan *social media* yang diproses juga sangat terbantu ketika dilakukan *parsing* dengan Python, karena cara dam *syntax* cukup *straightforward*.

2. Kode Mapper

```
#!/usr/bin/env python3
import json
import sys
from datetime import datetime
from dateutil import parser


def youtube_resource(specific_data):
    try:
        return specific_data["crawler_target"]["specific_resource_type"]
    except:
        return None
```

```
def twitter_resource(specific_data):
    try:
        return specific_data["crawler_target"]["specific_resource_type"]
    except:
        return None


def facebook_resource(specific_data):
    try:
        return specific_data["crawler_target"]["resource_type"]
    except:
        return None
def instagram_resource(specific_data):
    try:
        return specific_data["object"]["social_media"]
    except:
        return None


for line in sys.stdin:
    try:
        # Load the JSON object from the file
```

```
        data = json.loads(line)

        for obj in data:

            created_time = obj.get('created_time') or obj.get("created_at") or
obj.get("snippet").get("publishedAt") or
obj.get("snippet").get("topLevelComment").get("snippet").get("publishedAt")

            resource = youtube_resource(obj) or twitter_resource(obj) or
facebook_resource(obj) or instagram_resource(obj)

        # Extract the necessary data from the JSON object and emit key-value pairs

            if created_time.isdigit():

                created_time = datetime.fromtimestamp(int(created_time))

            else:

                created_time = parser.parse(created_time,yearfirst=True)

            if(created_time and resource):

                print(f'{created_time.date()}\t{resource}\t1')

    except:

        pass
```

Setiap file JSON memiliki struktur yang berbeda sedangkan dua properti yang digunakan dalam proses *mapping* hanya *social media* dan tanggal. Pertama-tama adalah melakukan abstraksi terhadap lokasi dari *social media* dan tanggal dari setiap kelompok masukan. Properti waktu memiliki sekitar 4 pattern, dan properti *social_media* juga memiliki sekitar 4

pattern. Ada sekelompok kecil masukan yang tidak terdeteksi, kebanyakan adalah yang berhubungan dengan kartu perdana (by.U, Telkomsel,dsb). Kelompok yang terdeteksi adalah Youtube, Twitter, Instagram, dan Facebook.

Setelah itu terdapat beberapa format tanggal yang akan dinormalisasi ke dalam bentuk YYYY-MM-DD.

Setelah itu hasil akan di print dengan separator tab, selanjutnya adalah bagian *reducer*.

3. Kode Reducer

```
#!/usr/bin/env python3
import sys
current_date = None
current_type = None
current_count = 0


for line in sys.stdin:
    date, soc_med_type, count = line.strip().split('\t')
    count = int(count)


    if date == current_date and soc_med_type == current_type:
        current_count += count
```

```
    else:
        if current_date and current_type:
            print(f'{current_date},{current_type},{current_count}')
        current_date = date
        current_type = soc_med_type
        current_count = count


if current_date and current_type:
    print(f'{current_date},{current_type},{current_count}')
```

4. Penjelasan alur MapReduce

Mapping dilakukan dengan proses parsing json dan mengambil data jenis sosial media (resource) dan tanggal pembuatan item sosial media tersebut (created_time). setelah parsing, mapping dilakukan dengan mengelompokan data berdasarkan created_time dan resource sosial media.

Kedua hal tersebut menjadi key agregat dengan value yaitu 1 sebagai jumlah kemunculannya. Lalu masuk kepada fungsi reducer, pada fungsi reducer ini akan menerima hasil keluaran mapper.

Pada fungsi reducer terjadi pengelompokkan berdasarkan created_time dan resource sosial media untuk dihitung jumlah

kemunculannya pada suatu tanggal dan suatu sosial media tertentu. Lalu data akan di print dalam bentuk csv.

5. Link github
   atlasfox007/Hadoop-Map-Reducer-IF4044 (github.com)