In [6]:
```
import sys
print(sys.version)

from pyspark.sql import SparkSession
spark = SparkSession.builder.getOrCreate()
spark.conf.set("spark.sql.repl.eagerEval.enabled",True)
print(spark.version)
```

```
3.8.15 | packaged by conda-forge | (default, Nov 22 2022, 08:46:39)
[GCC 10.4.0]
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLe
vel(newLevel).
22/12/09 02:33:11 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
22/12/09 02:33:11 INFO org.apache.spark.SparkEnv: Registering BlockManagerMast
er
22/12/09 02:33:11 INFO org.apache.spark.SparkEnv: Registering BlockManagerMast
erHeartbeat
22/12/09 02:33:11 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoor
dinator
3.1.3
```

In [7]:
```
import pandas as pd
import numpy as np
pd.set_option('display.max_colwidth', None)
pd.reset_option('display.max_rows')
from itertools import compress
from pyspark.sql.functions import *
from pyspark.sql.types import *
import seaborn as sns
import matplotlib.pyplot as plt
warnings.filterwarnings(action='ignore')
```

In [9]:
```
#Reading the twitter json file
twitter_df = spark.read.json('[redacted]')
```

```
22/12/09 02:48:30 WARN org.apache.spark.sql.execution.datasources.SharedInMemo
ryCache: Evicting cached table partition metadata from memory due to size cons
traints (spark.sql.hive.filesourcePartitionFileCacheSize = 262144000 bytes). T
his may impact query planning performance.
22/12/09 02:54:15 WARN org.apache.spark.sql.catalyst.util.package: Truncated t
he string representation of a plan since it was too large. This behavior can b
e adjusted by setting 'spark.sql.debug.maxToStringFields'.
```

In [19]:
```
twitter_df.limit(5)
```

Out[19]:

coordinates	created_at	display_text_range	entities	extended_entities	extended_tweet	favorit
null	Sun May 29 13:46:...	null	{[], null, [], []...	null	null	
null	Sun May 29 13:46:...	null	{[{[34, 43], SUNG...	null	null	
null	Sun May 29 13:46:...	null	{[], null, [], []...	null	null	
null	Sun May	null	{[], null,	null	null	

	29 13:46:...		[], []...			
null	Sun May 29 13:46:...	null	{[], null, [], []...	null	null	

In [116…
```
twitter_df.printSchema()
```

```
root
 |-- coordinates: struct (nullable = true)
 |    |-- coordinates: array (nullable = true)
 |    |    |-- element: double (containsNull = true)
 |    |-- type: string (nullable = true)
 |-- created_at: string (nullable = true)
 |-- display_text_range: array (nullable = true)
 |    |-- element: long (containsNull = true)
 |-- entities: struct (nullable = true)
 |    |-- hashtags: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- text: string (nullable = true)
 |    |-- media: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- additional_media_info: struct (nullable = true)
 |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |-- description: string (nullable = true)
 |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |-- id: long (nullable = true)
 |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |-- source_status_id_str: string (nullable = true)
 |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |-- type: string (nullable = true)
 |    |    |    |-- url: string (nullable = true)
 |    |-- symbols: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
```

```
 |    |    |    |-- text: string (nullable = true)
 |    |-- urls: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- url: string (nullable = true)
 |    |-- user_mentions: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- id: long (nullable = true)
 |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- name: string (nullable = true)
 |    |    |    |-- screen_name: string (nullable = true)
 |-- extended_entities: struct (nullable = true)
 |    |-- media: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- additional_media_info: struct (nullable = true)
 |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |-- description: string (nullable = true)
 |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |-- id: long (nullable = true)
 |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |-- source_status_id_str: string (nullable = true)
 |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |-- type: string (nullable = true)
 |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |-- content_type: string (nullable = true)
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |-- extended_tweet: struct (nullable = true)
 |    |-- display_text_range: array (nullable = true)
 |    |    |-- element: long (containsNull = true)
 |    |-- entities: struct (nullable = true)
```

```
|     |     |-- hashtags: array (nullable = true)
|     |     |    |-- element: struct (containsNull = true)
|     |     |    |    |-- indices: array (nullable = true)
|     |     |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |-- text: string (nullable = true)
|     |     |-- media: array (nullable = true)
|     |     |    |-- element: struct (containsNull = true)
|     |     |    |    |-- additional_media_info: struct (nullable = true)
|     |     |    |    |    |-- description: string (nullable = true)
|     |     |    |    |    |-- embeddable: boolean (nullable = true)
|     |     |    |    |    |-- monetizable: boolean (nullable = true)
|     |     |    |    |    |-- title: string (nullable = true)
|     |     |    |    |-- description: string (nullable = true)
|     |     |    |    |-- display_url: string (nullable = true)
|     |     |    |    |-- expanded_url: string (nullable = true)
|     |     |    |    |-- id: long (nullable = true)
|     |     |    |    |-- id_str: string (nullable = true)
|     |     |    |    |-- indices: array (nullable = true)
|     |     |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |-- media_url: string (nullable = true)
|     |     |    |    |-- media_url_https: string (nullable = true)
|     |     |    |    |-- sizes: struct (nullable = true)
|     |     |    |    |    |-- large: struct (nullable = true)
|     |     |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |-- medium: struct (nullable = true)
|     |     |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |-- small: struct (nullable = true)
|     |     |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |-- thumb: struct (nullable = true)
|     |     |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |-- source_status_id: long (nullable = true)
|     |     |    |    |-- source_status_id_str: string (nullable = true)
|     |     |    |    |-- source_user_id: long (nullable = true)
|     |     |    |    |-- source_user_id_str: string (nullable = true)
|     |     |    |    |-- type: string (nullable = true)
|     |     |    |    |-- url: string (nullable = true)
|     |     |    |    |-- video_info: struct (nullable = true)
|     |     |    |    |    |-- aspect_ratio: array (nullable = true)
|     |     |    |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |    |-- duration_millis: long (nullable = true)
|     |     |    |    |    |-- variants: array (nullable = true)
|     |     |    |    |    |    |-- element: struct (containsNull = true)
|     |     |    |    |    |    |    |-- bitrate: long (nullable = true)
|     |     |    |    |    |    |    |-- content_type: string (nullable = true)
|     |     |    |    |    |    |    |-- url: string (nullable = true)
|     |     |-- symbols: array (nullable = true)
|     |     |    |-- element: struct (containsNull = true)
|     |     |    |    |-- indices: array (nullable = true)
|     |     |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |-- text: string (nullable = true)
|     |     |-- urls: array (nullable = true)
|     |     |    |-- element: struct (containsNull = true)
|     |     |    |    |-- display_url: string (nullable = true)
|     |     |    |    |-- expanded_url: string (nullable = true)
|     |     |    |    |-- indices: array (nullable = true)
|     |     |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |-- url: string (nullable = true)
|     |     |-- user_mentions: array (nullable = true)
|     |     |    |-- element: struct (containsNull = true)
|     |     |    |    |-- id: long (nullable = true)
|     |     |    |    |-- id_str: string (nullable = true)
```

```
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |-- full_text: string (nullable = true)
|-- favorite_count: long (nullable = true)
|-- favorited: boolean (nullable = true)
|-- filter_level: string (nullable = true)
|-- geo: struct (nullable = true)
|    |-- coordinates: array (nullable = true)
|    |    |-- element: double (containsNull = true)
|    |-- type: string (nullable = true)
|-- id: long (nullable = true)
|-- id_str: string (nullable = true)
|-- in_reply_to_screen_name: string (nullable = true)
|-- in_reply_to_status_id: long (nullable = true)
|-- in_reply_to_status_id_str: string (nullable = true)
|-- in_reply_to_user_id: long (nullable = true)
|-- in_reply_to_user_id_str: string (nullable = true)
|-- is_quote_status: boolean (nullable = true)
```

```
|-- lang: string (nullable = true)
|-- place: struct (nullable = true)
|    |-- bounding_box: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: array (containsNull = true)
|    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- country: string (nullable = true)
|    |-- country_code: string (nullable = true)
|    |-- full_name: string (nullable = true)
|    |-- id: string (nullable = true)
|    |-- name: string (nullable = true)
|    |-- place_type: string (nullable = true)
|    |-- url: string (nullable = true)
|-- possibly_sensitive: boolean (nullable = true)
|-- quote_count: long (nullable = true)
|-- quoted_status: struct (nullable = true)
|    |-- coordinates: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- display_text_range: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
```

```
|     |     |     |     |-- url: string (nullable = true)
|     |     |     |-- symbols: array (nullable = true)
|     |     |     |    |-- element: struct (containsNull = true)
|     |     |     |    |    |-- indices: array (nullable = true)
|     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |    |    |-- text: string (nullable = true)
|     |     |     |-- urls: array (nullable = true)
|     |     |     |    |-- element: struct (containsNull = true)
|     |     |     |    |    |-- display_url: string (nullable = true)
|     |     |     |    |    |-- expanded_url: string (nullable = true)
|     |     |     |    |    |-- indices: array (nullable = true)
|     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |    |    |-- url: string (nullable = true)
|     |     |     |-- user_mentions: array (nullable = true)
|     |     |     |    |-- element: struct (containsNull = true)
|     |     |     |    |    |-- id: long (nullable = true)
|     |     |     |    |    |-- id_str: string (nullable = true)
|     |     |     |    |    |-- indices: array (nullable = true)
|     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |    |    |-- name: string (nullable = true)
|     |     |     |    |    |-- screen_name: string (nullable = true)
|     |     |-- extended_entities: struct (nullable = true)
|     |     |    |-- media: array (nullable = true)
|     |     |    |    |-- element: struct (containsNull = true)
|     |     |    |    |    |-- additional_media_info: struct (nullable = true)
|     |     |    |    |    |    |-- description: string (nullable = true)
|     |     |    |    |    |    |-- embeddable: boolean (nullable = true)
|     |     |    |    |    |    |-- monetizable: boolean (nullable = true)
|     |     |    |    |    |    |-- title: string (nullable = true)
|     |     |    |    |    |-- description: string (nullable = true)
|     |     |    |    |    |-- display_url: string (nullable = true)
|     |     |    |    |    |-- expanded_url: string (nullable = true)
|     |     |    |    |    |-- id: long (nullable = true)
|     |     |    |    |    |-- id_str: string (nullable = true)
|     |     |    |    |    |-- indices: array (nullable = true)
|     |     |    |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |    |-- media_url: string (nullable = true)
|     |     |    |    |    |-- media_url_https: string (nullable = true)
|     |     |    |    |    |-- sizes: struct (nullable = true)
|     |     |    |    |    |    |-- large: struct (nullable = true)
|     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |    |-- medium: struct (nullable = true)
|     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |    |-- small: struct (nullable = true)
|     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |    |-- thumb: struct (nullable = true)
|     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |    |    |    |-- source_status_id: long (nullable = true)
|     |     |    |    |    |-- source_status_id_str: string (nullable = true)
|     |     |    |    |    |-- source_user_id: long (nullable = true)
|     |     |    |    |    |-- source_user_id_str: string (nullable = true)
|     |     |    |    |    |-- type: string (nullable = true)
|     |     |    |    |    |-- url: string (nullable = true)
|     |     |    |    |    |-- video_info: struct (nullable = true)
|     |     |    |    |    |    |-- aspect_ratio: array (nullable = true)
|     |     |    |    |    |    |    |-- element: long (containsNull = true)
|     |     |    |    |    |    |-- duration_millis: long (nullable = true)
|     |     |    |    |    |    |-- variants: array (nullable = true)
|     |     |    |    |    |    |    |-- element: struct (containsNull = true)
|     |     |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|     |     |    |    |    |    |    |    |-- content_type: string (nullable = true)
```

```
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |-- extended_tweet: struct (nullable = true)
 |    |    |-- display_text_range: array (nullable = true)
 |    |    |    |-- element: long (containsNull = true)
 |    |    |-- entities: struct (nullable = true)
 |    |    |    |-- hashtags: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |-- media: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
 |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
 |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- symbols: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |-- urls: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
```

```
 |     |     |     |     |     |    |-- element: long (containsNull = true)
 |     |     |     |     |     |-- url: string (nullable = true)
 |     |     |     |-- user_mentions: array (nullable = true)
 |     |     |     |    |-- element: struct (containsNull = true)
 |     |     |     |    |    |-- id: long (nullable = true)
 |     |     |     |    |    |-- id_str: string (nullable = true)
 |     |     |     |    |    |-- indices: array (nullable = true)
 |     |     |     |    |    |    |-- element: long (containsNull = true)
 |     |     |     |    |    |-- name: string (nullable = true)
 |     |     |     |    |    |-- screen_name: string (nullable = true)
 |     |     |-- extended_entities: struct (nullable = true)
 |     |     |    |-- media: array (nullable = true)
 |     |     |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |-- additional_media_info: struct (nullable = true)
 |     |     |    |    |    |    |-- description: string (nullable = true)
 |     |     |    |    |    |    |-- embeddable: boolean (nullable = true)
 |     |     |    |    |    |    |-- monetizable: boolean (nullable = true)
 |     |     |    |    |    |    |-- title: string (nullable = true)
 |     |     |    |    |    |-- description: string (nullable = true)
 |     |     |    |    |    |-- display_url: string (nullable = true)
 |     |     |    |    |    |-- expanded_url: string (nullable = true)
 |     |     |    |    |    |-- id: long (nullable = true)
 |     |     |    |    |    |-- id_str: string (nullable = true)
 |     |     |    |    |    |-- indices: array (nullable = true)
 |     |     |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |-- media_url: string (nullable = true)
 |     |     |    |    |    |-- media_url_https: string (nullable = true)
 |     |     |    |    |    |-- sizes: struct (nullable = true)
 |     |     |    |    |    |    |-- large: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- medium: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- small: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- thumb: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |-- source_status_id: long (nullable = true)
 |     |     |    |    |    |-- source_status_id_str: string (nullable = true)
 |     |     |    |    |    |-- source_user_id: long (nullable = true)
 |     |     |    |    |    |-- source_user_id_str: string (nullable = true)
 |     |     |    |    |    |-- type: string (nullable = true)
 |     |     |    |    |    |-- url: string (nullable = true)
 |     |     |    |    |    |-- video_info: struct (nullable = true)
 |     |     |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |     |     |    |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |    |-- duration_millis: long (nullable = true)
 |     |     |    |    |    |    |-- variants: array (nullable = true)
 |     |     |    |    |    |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |     |     |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |     |     |    |    |    |    |    |    |-- url: string (nullable = true)
 |     |     |    |-- full_text: string (nullable = true)
 |     |-- favorite_count: long (nullable = true)
 |     |-- favorited: boolean (nullable = true)
 |     |-- filter_level: string (nullable = true)
 |     |-- geo: struct (nullable = true)
 |     |    |-- coordinates: array (nullable = true)
 |     |    |    |-- element: double (containsNull = true)
 |     |    |-- type: string (nullable = true)
 |     |-- id: long (nullable = true)
```

```
|    |-- id_str: string (nullable = true)
|    |-- in_reply_to_screen_name: string (nullable = true)
|    |-- in_reply_to_status_id: long (nullable = true)
|    |-- in_reply_to_status_id_str: string (nullable = true)
|    |-- in_reply_to_user_id: long (nullable = true)
|    |-- in_reply_to_user_id_str: string (nullable = true)
|    |-- is_quote_status: boolean (nullable = true)
|    |-- lang: string (nullable = true)
|    |-- place: struct (nullable = true)
|    |    |-- bounding_box: struct (nullable = true)
|    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |    |-- element: double (containsNull = true)
|    |    |    |-- type: string (nullable = true)
|    |    |-- country: string (nullable = true)
|    |    |-- country_code: string (nullable = true)
|    |    |-- full_name: string (nullable = true)
|    |    |-- id: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- place_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |-- possibly_sensitive: boolean (nullable = true)
|    |-- quote_count: long (nullable = true)
|    |-- quoted_status_id: long (nullable = true)
|    |-- quoted_status_id_str: string (nullable = true)
|    |-- reply_count: long (nullable = true)
|    |-- retweet_count: long (nullable = true)
|    |-- retweeted: boolean (nullable = true)
|    |-- scopes: struct (nullable = true)
|    |    |-- followers: boolean (nullable = true)
|    |-- source: string (nullable = true)
|    |-- text: string (nullable = true)
|    |-- truncated: boolean (nullable = true)
|    |-- user: struct (nullable = true)
|    |    |-- contributors_enabled: boolean (nullable = true)
|    |    |-- created_at: string (nullable = true)
|    |    |-- default_profile: boolean (nullable = true)
|    |    |-- default_profile_image: boolean (nullable = true)
|    |    |-- description: string (nullable = true)
|    |    |-- favourites_count: long (nullable = true)
|    |    |-- followers_count: long (nullable = true)
|    |    |-- friends_count: long (nullable = true)
|    |    |-- geo_enabled: boolean (nullable = true)
|    |    |-- id: long (nullable = true)
|    |    |-- id_str: string (nullable = true)
|    |    |-- is_translator: boolean (nullable = true)
|    |    |-- listed_count: long (nullable = true)
|    |    |-- location: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- profile_background_color: string (nullable = true)
|    |    |-- profile_background_image_url: string (nullable = true)
|    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |-- profile_banner_url: string (nullable = true)
|    |    |-- profile_image_url: string (nullable = true)
|    |    |-- profile_image_url_https: string (nullable = true)
|    |    |-- profile_link_color: string (nullable = true)
|    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |-- profile_text_color: string (nullable = true)
|    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |-- protected: boolean (nullable = true)
|    |    |-- screen_name: string (nullable = true)
|    |    |-- statuses_count: long (nullable = true)
|    |    |-- translator_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |    |-- verified: boolean (nullable = true)
|    |    |-- withheld_in_countries: array (nullable = true)
```

```
|     |     |     |-- element: string (containsNull = true)
|     |-- withheld_copyright: boolean (nullable = true)
|     |-- withheld_in_countries: array (nullable = true)
|     |     |-- element: string (containsNull = true)
|-- quoted_status_id: long (nullable = true)
|-- quoted_status_id_str: string (nullable = true)
|-- quoted_status_permalink: struct (nullable = true)
|     |-- display: string (nullable = true)
|     |-- expanded: string (nullable = true)
|     |-- url: string (nullable = true)
|-- quoted_text: string (nullable = true)
|-- reply_count: long (nullable = true)
|-- retweet_count: long (nullable = true)
|-- retweeted: string (nullable = true)
|-- retweeted_from: string (nullable = true)
|-- retweeted_status: struct (nullable = true)
|     |-- coordinates: struct (nullable = true)
|     |     |-- coordinates: array (nullable = true)
|     |     |     |-- element: double (containsNull = true)
|     |     |-- type: string (nullable = true)
|     |-- created_at: string (nullable = true)
|     |-- display_text_range: array (nullable = true)
|     |     |-- element: long (containsNull = true)
|     |-- entities: struct (nullable = true)
|     |     |-- hashtags: array (nullable = true)
|     |     |     |-- element: struct (containsNull = true)
|     |     |     |     |-- indices: array (nullable = true)
|     |     |     |     |     |-- element: long (containsNull = true)
|     |     |     |     |-- text: string (nullable = true)
|     |     |-- media: array (nullable = true)
|     |     |     |-- element: struct (containsNull = true)
|     |     |     |     |-- additional_media_info: struct (nullable = true)
|     |     |     |     |     |-- description: string (nullable = true)
|     |     |     |     |     |-- embeddable: boolean (nullable = true)
|     |     |     |     |     |-- monetizable: boolean (nullable = true)
|     |     |     |     |     |-- title: string (nullable = true)
|     |     |     |     |-- description: string (nullable = true)
|     |     |     |     |-- display_url: string (nullable = true)
|     |     |     |     |-- expanded_url: string (nullable = true)
|     |     |     |     |-- id: long (nullable = true)
|     |     |     |     |-- id_str: string (nullable = true)
|     |     |     |     |-- indices: array (nullable = true)
|     |     |     |     |     |-- element: long (containsNull = true)
|     |     |     |     |-- media_url: string (nullable = true)
|     |     |     |     |-- media_url_https: string (nullable = true)
|     |     |     |     |-- sizes: struct (nullable = true)
|     |     |     |     |     |-- large: struct (nullable = true)
|     |     |     |     |     |     |-- h: long (nullable = true)
|     |     |     |     |     |     |-- resize: string (nullable = true)
|     |     |     |     |     |     |-- w: long (nullable = true)
|     |     |     |     |     |-- medium: struct (nullable = true)
|     |     |     |     |     |     |-- h: long (nullable = true)
|     |     |     |     |     |     |-- resize: string (nullable = true)
|     |     |     |     |     |     |-- w: long (nullable = true)
|     |     |     |     |     |-- small: struct (nullable = true)
|     |     |     |     |     |     |-- h: long (nullable = true)
|     |     |     |     |     |     |-- resize: string (nullable = true)
|     |     |     |     |     |     |-- w: long (nullable = true)
|     |     |     |     |     |-- thumb: struct (nullable = true)
|     |     |     |     |     |     |-- h: long (nullable = true)
|     |     |     |     |     |     |-- resize: string (nullable = true)
|     |     |     |     |     |     |-- w: long (nullable = true)
|     |     |     |     |-- source_status_id: long (nullable = true)
|     |     |     |     |-- source_status_id_str: string (nullable = true)
|     |     |     |     |-- source_user_id: long (nullable = true)
|     |     |     |     |-- source_user_id_str: string (nullable = true)
|     |     |     |     |-- type: string (nullable = true)
|     |     |     |     |-- url: string (nullable = true)
|     |     |-- symbols: array (nullable = true)
```

```
|     |     |     |-- element: struct (containsNull = true)
|     |     |     |    |-- indices: array (nullable = true)
|     |     |     |    |    |-- element: long (containsNull = true)
|     |     |     |    |-- text: string (nullable = true)
|     |     |-- urls: array (nullable = true)
|     |     |     |-- element: struct (containsNull = true)
|     |     |     |    |-- display_url: string (nullable = true)
|     |     |     |    |-- expanded_url: string (nullable = true)
|     |     |     |    |-- indices: array (nullable = true)
|     |     |     |    |    |-- element: long (containsNull = true)
|     |     |     |    |-- url: string (nullable = true)
|     |     |-- user_mentions: array (nullable = true)
|     |     |     |-- element: struct (containsNull = true)
|     |     |     |    |-- id: long (nullable = true)
|     |     |     |    |-- id_str: string (nullable = true)
|     |     |     |    |-- indices: array (nullable = true)
|     |     |     |    |    |-- element: long (containsNull = true)
|     |     |     |    |-- name: string (nullable = true)
|     |     |     |    |-- screen_name: string (nullable = true)
|     |-- extended_entities: struct (nullable = true)
|     |     |-- media: array (nullable = true)
|     |     |     |-- element: struct (containsNull = true)
|     |     |     |    |-- additional_media_info: struct (nullable = true)
|     |     |     |    |    |-- description: string (nullable = true)
|     |     |     |    |    |-- embeddable: boolean (nullable = true)
|     |     |     |    |    |-- monetizable: boolean (nullable = true)
|     |     |     |    |    |-- title: string (nullable = true)
|     |     |     |    |-- description: string (nullable = true)
|     |     |     |    |-- display_url: string (nullable = true)
|     |     |     |    |-- expanded_url: string (nullable = true)
|     |     |     |    |-- id: long (nullable = true)
|     |     |     |    |-- id_str: string (nullable = true)
|     |     |     |    |-- indices: array (nullable = true)
|     |     |     |    |    |-- element: long (containsNull = true)
|     |     |     |    |-- media_url: string (nullable = true)
|     |     |     |    |-- media_url_https: string (nullable = true)
|     |     |     |    |-- sizes: struct (nullable = true)
|     |     |     |    |    |-- large: struct (nullable = true)
|     |     |     |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |-- medium: struct (nullable = true)
|     |     |     |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |-- small: struct (nullable = true)
|     |     |     |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |-- thumb: struct (nullable = true)
|     |     |     |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |-- w: long (nullable = true)
|     |     |     |    |-- source_status_id: long (nullable = true)
|     |     |     |    |-- source_status_id_str: string (nullable = true)
|     |     |     |    |-- source_user_id: long (nullable = true)
|     |     |     |    |-- source_user_id_str: string (nullable = true)
|     |     |     |    |-- type: string (nullable = true)
|     |     |     |    |-- url: string (nullable = true)
|     |     |     |    |-- video_info: struct (nullable = true)
|     |     |     |    |    |-- aspect_ratio: array (nullable = true)
|     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |    |    |-- duration_millis: long (nullable = true)
|     |     |     |    |    |-- variants: array (nullable = true)
|     |     |     |    |    |    |-- element: struct (containsNull = true)
|     |     |     |    |    |    |    |-- bitrate: long (nullable = true)
|     |     |     |    |    |    |    |-- content_type: string (nullable = true)
|     |     |     |    |    |    |    |-- url: string (nullable = true)
|     |-- extended_tweet: struct (nullable = true)
```

```
 |     |     |-- display_text_range: array (nullable = true)
 |     |     |    |-- element: long (containsNull = true)
 |     |     |-- entities: struct (nullable = true)
 |     |     |    |-- hashtags: array (nullable = true)
 |     |     |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |-- indices: array (nullable = true)
 |     |     |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |-- text: string (nullable = true)
 |     |     |    |-- media: array (nullable = true)
 |     |     |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |-- additional_media_info: struct (nullable = true)
 |     |     |    |    |    |    |-- description: string (nullable = true)
 |     |     |    |    |    |    |-- embeddable: boolean (nullable = true)
 |     |     |    |    |    |    |-- monetizable: boolean (nullable = true)
 |     |     |    |    |    |    |-- title: string (nullable = true)
 |     |     |    |    |    |-- description: string (nullable = true)
 |     |     |    |    |    |-- display_url: string (nullable = true)
 |     |     |    |    |    |-- expanded_url: string (nullable = true)
 |     |     |    |    |    |-- id: long (nullable = true)
 |     |     |    |    |    |-- id_str: string (nullable = true)
 |     |     |    |    |    |-- indices: array (nullable = true)
 |     |     |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |-- media_url: string (nullable = true)
 |     |     |    |    |    |-- media_url_https: string (nullable = true)
 |     |     |    |    |    |-- sizes: struct (nullable = true)
 |     |     |    |    |    |    |-- large: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- medium: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- small: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- thumb: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |-- source_status_id: long (nullable = true)
 |     |     |    |    |    |-- source_status_id_str: string (nullable = true)
 |     |     |    |    |    |-- source_user_id: long (nullable = true)
 |     |     |    |    |    |-- source_user_id_str: string (nullable = true)
 |     |     |    |    |    |-- type: string (nullable = true)
 |     |     |    |    |    |-- url: string (nullable = true)
 |     |     |    |    |    |-- video_info: struct (nullable = true)
 |     |     |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |     |     |    |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |    |-- duration_millis: long (nullable = true)
 |     |     |    |    |    |    |-- variants: array (nullable = true)
 |     |     |    |    |    |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |     |     |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |     |     |    |    |    |    |    |    |-- url: string (nullable = true)
 |     |     |    |-- symbols: array (nullable = true)
 |     |     |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |-- indices: array (nullable = true)
 |     |     |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |-- text: string (nullable = true)
 |     |     |    |-- urls: array (nullable = true)
 |     |     |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |-- display_url: string (nullable = true)
 |     |     |    |    |    |-- expanded_url: string (nullable = true)
 |     |     |    |    |    |-- indices: array (nullable = true)
 |     |     |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |-- url: string (nullable = true)
```

```
 |     |     |     |-- user_mentions: array (nullable = true)
 |     |     |     |    |-- element: struct (containsNull = true)
 |     |     |     |    |    |-- id: long (nullable = true)
 |     |     |     |    |    |-- id_str: string (nullable = true)
 |     |     |     |    |    |-- indices: array (nullable = true)
 |     |     |     |    |    |    |-- element: long (containsNull = true)
 |     |     |     |    |    |-- name: string (nullable = true)
 |     |     |     |    |    |-- screen_name: string (nullable = true)
 |     |     |-- extended_entities: struct (nullable = true)
 |     |     |    |-- media: array (nullable = true)
 |     |     |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |-- additional_media_info: struct (nullable = true)
 |     |     |    |    |    |    |-- description: string (nullable = true)
 |     |     |    |    |    |    |-- embeddable: boolean (nullable = true)
 |     |     |    |    |    |    |-- monetizable: boolean (nullable = true)
 |     |     |    |    |    |    |-- title: string (nullable = true)
 |     |     |    |    |    |-- description: string (nullable = true)
 |     |     |    |    |    |-- display_url: string (nullable = true)
 |     |     |    |    |    |-- expanded_url: string (nullable = true)
 |     |     |    |    |    |-- id: long (nullable = true)
 |     |     |    |    |    |-- id_str: string (nullable = true)
 |     |     |    |    |    |-- indices: array (nullable = true)
 |     |     |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |-- media_url: string (nullable = true)
 |     |     |    |    |    |-- media_url_https: string (nullable = true)
 |     |     |    |    |    |-- sizes: struct (nullable = true)
 |     |     |    |    |    |    |-- large: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- medium: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- small: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |    |-- thumb: struct (nullable = true)
 |     |     |    |    |    |    |    |-- h: long (nullable = true)
 |     |     |    |    |    |    |    |-- resize: string (nullable = true)
 |     |     |    |    |    |    |    |-- w: long (nullable = true)
 |     |     |    |    |    |-- source_status_id: long (nullable = true)
 |     |     |    |    |    |-- source_status_id_str: string (nullable = true)
 |     |     |    |    |    |-- source_user_id: long (nullable = true)
 |     |     |    |    |    |-- source_user_id_str: string (nullable = true)
 |     |     |    |    |    |-- type: string (nullable = true)
 |     |     |    |    |    |-- url: string (nullable = true)
 |     |     |    |    |    |-- video_info: struct (nullable = true)
 |     |     |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |     |     |    |    |    |    |    |-- element: long (containsNull = true)
 |     |     |    |    |    |    |-- duration_millis: long (nullable = true)
 |     |     |    |    |    |    |-- variants: array (nullable = true)
 |     |     |    |    |    |    |    |-- element: struct (containsNull = true)
 |     |     |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |     |     |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |     |     |    |    |    |    |    |    |-- url: string (nullable = true)
 |     |     |    |-- full_text: string (nullable = true)
 |     |-- favorite_count: long (nullable = true)
 |     |-- favorited: boolean (nullable = true)
 |     |-- filter_level: string (nullable = true)
 |     |-- geo: struct (nullable = true)
 |     |    |-- coordinates: array (nullable = true)
 |     |    |    |-- element: double (containsNull = true)
 |     |    |-- type: string (nullable = true)
 |     |-- id: long (nullable = true)
 |     |-- id_str: string (nullable = true)
 |     |-- in_reply_to_screen_name: string (nullable = true)
```

```
|     |-- in_reply_to_status_id: long (nullable = true)
|     |-- in_reply_to_status_id_str: string (nullable = true)
|     |-- in_reply_to_user_id: long (nullable = true)
|     |-- in_reply_to_user_id_str: string (nullable = true)
|     |-- is_quote_status: boolean (nullable = true)
|     |-- lang: string (nullable = true)
|     |-- place: struct (nullable = true)
|     |    |-- bounding_box: struct (nullable = true)
|     |    |    |-- coordinates: array (nullable = true)
|     |    |    |    |-- element: array (containsNull = true)
|     |    |    |    |    |-- element: array (containsNull = true)
|     |    |    |    |    |    |-- element: double (containsNull = true)
|     |    |    |-- type: string (nullable = true)
|     |    |-- country: string (nullable = true)
|     |    |-- country_code: string (nullable = true)
|     |    |-- full_name: string (nullable = true)
|     |    |-- id: string (nullable = true)
|     |    |-- name: string (nullable = true)
|     |    |-- place_type: string (nullable = true)
|     |    |-- url: string (nullable = true)
|     |-- possibly_sensitive: boolean (nullable = true)
|     |-- quote_count: long (nullable = true)
|     |-- quoted_status: struct (nullable = true)
|     |    |-- coordinates: struct (nullable = true)
|     |    |    |-- coordinates: array (nullable = true)
|     |    |    |    |-- element: double (containsNull = true)
|     |    |    |-- type: string (nullable = true)
|     |    |-- created_at: string (nullable = true)
|     |    |-- display_text_range: array (nullable = true)
|     |    |    |-- element: long (containsNull = true)
|     |    |-- entities: struct (nullable = true)
|     |    |    |-- hashtags: array (nullable = true)
|     |    |    |    |-- element: struct (containsNull = true)
|     |    |    |    |    |-- indices: array (nullable = true)
|     |    |    |    |    |    |-- element: long (containsNull = true)
|     |    |    |    |    |-- text: string (nullable = true)
|     |    |    |-- media: array (nullable = true)
|     |    |    |    |-- element: struct (containsNull = true)
|     |    |    |    |    |-- additional_media_info: struct (nullable = true)
|     |    |    |    |    |    |-- description: string (nullable = true)
|     |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|     |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|     |    |    |    |    |    |-- title: string (nullable = true)
|     |    |    |    |    |-- description: string (nullable = true)
|     |    |    |    |    |-- display_url: string (nullable = true)
|     |    |    |    |    |-- expanded_url: string (nullable = true)
|     |    |    |    |    |-- id: long (nullable = true)
|     |    |    |    |    |-- id_str: string (nullable = true)
|     |    |    |    |    |-- indices: array (nullable = true)
|     |    |    |    |    |    |-- element: long (containsNull = true)
|     |    |    |    |    |-- media_url: string (nullable = true)
|     |    |    |    |    |-- media_url_https: string (nullable = true)
|     |    |    |    |    |-- sizes: struct (nullable = true)
|     |    |    |    |    |    |-- large: struct (nullable = true)
|     |    |    |    |    |    |    |-- h: long (nullable = true)
|     |    |    |    |    |    |    |-- resize: string (nullable = true)
|     |    |    |    |    |    |    |-- w: long (nullable = true)
|     |    |    |    |    |    |-- medium: struct (nullable = true)
|     |    |    |    |    |    |    |-- h: long (nullable = true)
|     |    |    |    |    |    |    |-- resize: string (nullable = true)
|     |    |    |    |    |    |    |-- w: long (nullable = true)
|     |    |    |    |    |    |-- small: struct (nullable = true)
|     |    |    |    |    |    |    |-- h: long (nullable = true)
|     |    |    |    |    |    |    |-- resize: string (nullable = true)
|     |    |    |    |    |    |    |-- w: long (nullable = true)
|     |    |    |    |    |    |-- thumb: struct (nullable = true)
|     |    |    |    |    |    |    |-- h: long (nullable = true)
|     |    |    |    |    |    |    |-- resize: string (nullable = true)
|     |    |    |    |    |    |    |-- w: long (nullable = true)
```

```
|     |     |     |     |     |-- source_status_id: long (nullable = true)
|     |     |     |     |     |-- source_status_id_str: string (nullable = true)
|     |     |     |     |     |-- source_user_id: long (nullable = true)
|     |     |     |     |     |-- source_user_id_str: string (nullable = true)
|     |     |     |     |     |-- type: string (nullable = true)
|     |     |     |     |     |-- url: string (nullable = true)
|     |     |     |     |-- symbols: array (nullable = true)
|     |     |     |     |    |-- element: struct (containsNull = true)
|     |     |     |     |    |    |-- indices: array (nullable = true)
|     |     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |     |    |    |-- text: string (nullable = true)
|     |     |     |     |-- urls: array (nullable = true)
|     |     |     |     |    |-- element: struct (containsNull = true)
|     |     |     |     |    |    |-- display_url: string (nullable = true)
|     |     |     |     |    |    |-- expanded_url: string (nullable = true)
|     |     |     |     |    |    |-- indices: array (nullable = true)
|     |     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |     |    |    |-- url: string (nullable = true)
|     |     |     |     |-- user_mentions: array (nullable = true)
|     |     |     |     |    |-- element: struct (containsNull = true)
|     |     |     |     |    |    |-- id: long (nullable = true)
|     |     |     |     |    |    |-- id_str: string (nullable = true)
|     |     |     |     |    |    |-- indices: array (nullable = true)
|     |     |     |     |    |    |    |-- element: long (containsNull = true)
|     |     |     |     |    |    |-- name: string (nullable = true)
|     |     |     |     |    |    |-- screen_name: string (nullable = true)
|     |     |     |-- extended_entities: struct (nullable = true)
|     |     |     |    |-- media: array (nullable = true)
|     |     |     |    |    |-- element: struct (containsNull = true)
|     |     |     |    |    |    |-- additional_media_info: struct (nullable = true)
|     |     |     |    |    |    |    |-- description: string (nullable = true)
|     |     |     |    |    |    |    |-- embeddable: boolean (nullable = true)
|     |     |     |    |    |    |    |-- monetizable: boolean (nullable = true)
|     |     |     |    |    |    |    |-- title: string (nullable = true)
|     |     |     |    |    |    |-- description: string (nullable = true)
|     |     |     |    |    |    |-- display_url: string (nullable = true)
|     |     |     |    |    |    |-- expanded_url: string (nullable = true)
|     |     |     |    |    |    |-- id: long (nullable = true)
|     |     |     |    |    |    |-- id_str: string (nullable = true)
|     |     |     |    |    |    |-- indices: array (nullable = true)
|     |     |     |    |    |    |    |-- element: long (containsNull = true)
|     |     |     |    |    |    |-- media_url: string (nullable = true)
|     |     |     |    |    |    |-- media_url_https: string (nullable = true)
|     |     |     |    |    |    |-- sizes: struct (nullable = true)
|     |     |     |    |    |    |    |-- large: struct (nullable = true)
|     |     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |    |    |-- medium: struct (nullable = true)
|     |     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |    |    |-- small: struct (nullable = true)
|     |     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |    |    |-- thumb: struct (nullable = true)
|     |     |     |    |    |    |    |    |-- h: long (nullable = true)
|     |     |     |    |    |    |    |    |-- resize: string (nullable = true)
|     |     |     |    |    |    |    |    |-- w: long (nullable = true)
|     |     |     |    |    |    |-- source_status_id: long (nullable = true)
|     |     |     |    |    |    |-- source_status_id_str: string (nullable = true)
|     |     |     |    |    |    |-- source_user_id: long (nullable = true)
|     |     |     |    |    |    |-- source_user_id_str: string (nullable = true)
|     |     |     |    |    |    |-- type: string (nullable = true)
|     |     |     |    |    |    |-- url: string (nullable = true)
|     |     |     |    |    |    |-- video_info: struct (nullable = true)
|     |     |     |    |    |    |    |-- aspect_ratio: array (nullable = true)
|     |     |     |    |    |    |    |    |-- element: long (containsNull = true)
```

```
 |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |-- extended_tweet: struct (nullable = true)
 |    |    |    |-- display_text_range: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- entities: struct (nullable = true)
 |    |    |    |    |-- hashtags: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |    |-- media: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- additional_media_info: struct (nullable = t
rue)
 |    |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id_str: string (nullable = tr
ue)
 |    |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_user_id_str: string (nullable = tru
e)
 |    |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: long (containsNull = tru
e)
 |    |    |    |    |    |    |    |-- duration_millis: long (nullable = tru
e)
 |    |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: struct (containsNull = t
rue)
 |    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = tr
ue)
```

```
 |    |    |    |    |    |    |    |    |-- content_type: string (nullab
le = true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = tru
e)
 |    |    |    |    |-- symbols: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |    |-- urls: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |-- user_mentions: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- name: string (nullable = true)
 |    |    |    |    |    |    |-- screen_name: string (nullable = true)
 |    |    |    |-- extended_entities: struct (nullable = true)
 |    |    |    |    |-- media: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- additional_media_info: struct (nullable = t
rue)
 |    |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id_str: string (nullable = tr
ue)
 |    |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_user_id_str: string (nullable = tru
e)
 |    |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
```

```
 |    |    |    |    |    |    |    |    |-- element: long (containsNull = tru
e)
 |    |    |    |    |    |    |    |-- duration_millis: long (nullable = tru
e)
 |    |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: struct (containsNull = t
rue)
 |    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = tr
ue)
 |    |    |    |    |    |    |    |    |    |-- content_type: string (nullab
le = true)
 |    |    |    |    |    |    |    |    |    |-- url: string (nullable = tru
e)
 |    |    |    |-- full_text: string (nullable = true)
 |    |    |-- favorite_count: long (nullable = true)
 |    |    |-- favorited: boolean (nullable = true)
 |    |    |-- filter_level: string (nullable = true)
 |    |    |-- geo: struct (nullable = true)
 |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |-- element: double (containsNull = true)
 |    |    |    |-- type: string (nullable = true)
 |    |    |-- id: long (nullable = true)
 |    |    |-- id_str: string (nullable = true)
 |    |    |-- in_reply_to_screen_name: string (nullable = true)
 |    |    |-- in_reply_to_status_id: long (nullable = true)
 |    |    |-- in_reply_to_status_id_str: string (nullable = true)
 |    |    |-- in_reply_to_user_id: long (nullable = true)
 |    |    |-- in_reply_to_user_id_str: string (nullable = true)
 |    |    |-- is_quote_status: boolean (nullable = true)
 |    |    |-- lang: string (nullable = true)
 |    |    |-- place: struct (nullable = true)
 |    |    |    |-- bounding_box: struct (nullable = true)
 |    |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |    |    |-- element: double (containsNull = true)
 |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |-- country: string (nullable = true)
 |    |    |    |-- country_code: string (nullable = true)
 |    |    |    |-- full_name: string (nullable = true)
 |    |    |    |-- id: string (nullable = true)
 |    |    |    |-- name: string (nullable = true)
 |    |    |    |-- place_type: string (nullable = true)
 |    |    |    |-- url: string (nullable = true)
 |    |    |-- possibly_sensitive: boolean (nullable = true)
 |    |    |-- quote_count: long (nullable = true)
 |    |    |-- quoted_status_id: long (nullable = true)
 |    |    |-- quoted_status_id_str: string (nullable = true)
 |    |    |-- reply_count: long (nullable = true)
 |    |    |-- retweet_count: long (nullable = true)
 |    |    |-- retweeted: boolean (nullable = true)
 |    |    |-- scopes: struct (nullable = true)
 |    |    |    |-- followers: boolean (nullable = true)
 |    |    |-- source: string (nullable = true)
 |    |    |-- text: string (nullable = true)
 |    |    |-- truncated: boolean (nullable = true)
 |    |    |-- user: struct (nullable = true)
 |    |    |    |-- contributors_enabled: boolean (nullable = true)
 |    |    |    |-- created_at: string (nullable = true)
 |    |    |    |-- default_profile: boolean (nullable = true)
 |    |    |    |-- default_profile_image: boolean (nullable = true)
 |    |    |    |-- description: string (nullable = true)
 |    |    |    |-- favourites_count: long (nullable = true)
 |    |    |    |-- followers_count: long (nullable = true)
 |    |    |    |-- friends_count: long (nullable = true)
 |    |    |    |-- geo_enabled: boolean (nullable = true)
 |    |    |    |-- id: long (nullable = true)
 |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |-- is_translator: boolean (nullable = true)
```

```
|     |     |     |-- listed_count: long (nullable = true)
|     |     |     |-- location: string (nullable = true)
|     |     |     |-- name: string (nullable = true)
|     |     |     |-- profile_background_color: string (nullable = true)
|     |     |     |-- profile_background_image_url: string (nullable = true)
|     |     |     |-- profile_background_image_url_https: string (nullable = tru
e)
|     |     |     |-- profile_background_tile: boolean (nullable = true)
|     |     |     |-- profile_banner_url: string (nullable = true)
|     |     |     |-- profile_image_url: string (nullable = true)
|     |     |     |-- profile_image_url_https: string (nullable = true)
|     |     |     |-- profile_link_color: string (nullable = true)
|     |     |     |-- profile_sidebar_border_color: string (nullable = true)
|     |     |     |-- profile_sidebar_fill_color: string (nullable = true)
|     |     |     |-- profile_text_color: string (nullable = true)
|     |     |     |-- profile_use_background_image: boolean (nullable = true)
|     |     |     |-- protected: boolean (nullable = true)
|     |     |     |-- screen_name: string (nullable = true)
|     |     |     |-- statuses_count: long (nullable = true)
|     |     |     |-- translator_type: string (nullable = true)
|     |     |     |-- url: string (nullable = true)
|     |     |     |-- verified: boolean (nullable = true)
|     |     |     |-- withheld_in_countries: array (nullable = true)
|     |     |     |    |-- element: string (containsNull = true)
|     |     |-- withheld_in_countries: array (nullable = true)
|     |     |    |-- element: string (containsNull = true)
|     |-- quoted_status_id: long (nullable = true)
|     |-- quoted_status_id_str: string (nullable = true)
|     |-- quoted_status_permalink: struct (nullable = true)
|     |    |-- display: string (nullable = true)
|     |    |-- expanded: string (nullable = true)
|     |    |-- url: string (nullable = true)
|     |-- reply_count: long (nullable = true)
|     |-- retweet_count: long (nullable = true)
|     |-- retweeted: boolean (nullable = true)
|     |-- scopes: struct (nullable = true)
|     |    |-- followers: boolean (nullable = true)
|     |    |-- place_ids: array (nullable = true)
|     |    |    |-- element: string (containsNull = true)
|     |-- source: string (nullable = true)
|     |-- text: string (nullable = true)
|     |-- truncated: boolean (nullable = true)
|     |-- user: struct (nullable = true)
|     |    |-- contributors_enabled: boolean (nullable = true)
|     |    |-- created_at: string (nullable = true)
|     |    |-- default_profile: boolean (nullable = true)
|     |    |-- default_profile_image: boolean (nullable = true)
|     |    |-- description: string (nullable = true)
|     |    |-- favourites_count: long (nullable = true)
|     |    |-- followers_count: long (nullable = true)
|     |    |-- friends_count: long (nullable = true)
|     |    |-- geo_enabled: boolean (nullable = true)
|     |    |-- id: long (nullable = true)
|     |    |-- id_str: string (nullable = true)
|     |    |-- is_translator: boolean (nullable = true)
|     |    |-- listed_count: long (nullable = true)
|     |    |-- location: string (nullable = true)
|     |    |-- name: string (nullable = true)
|     |    |-- profile_background_color: string (nullable = true)
|     |    |-- profile_background_image_url: string (nullable = true)
|     |    |-- profile_background_image_url_https: string (nullable = true)
|     |    |-- profile_background_tile: boolean (nullable = true)
|     |    |-- profile_banner_url: string (nullable = true)
|     |    |-- profile_image_url: string (nullable = true)
|     |    |-- profile_image_url_https: string (nullable = true)
|     |    |-- profile_link_color: string (nullable = true)
|     |    |-- profile_sidebar_border_color: string (nullable = true)
|     |    |-- profile_sidebar_fill_color: string (nullable = true)
|     |    |-- profile_text_color: string (nullable = true)
```

```
|     |     |-- profile_use_background_image: boolean (nullable = true)
|     |     |-- protected: boolean (nullable = true)
|     |     |-- screen_name: string (nullable = true)
|     |     |-- statuses_count: long (nullable = true)
|     |     |-- translator_type: string (nullable = true)
|     |     |-- url: string (nullable = true)
|     |     |-- verified: boolean (nullable = true)
|     |     |-- withheld_in_countries: array (nullable = true)
|     |     |     |-- element: string (containsNull = true)
|     |-- withheld_copyright: boolean (nullable = true)
|     |-- withheld_in_countries: array (nullable = true)
|     |     |-- element: string (containsNull = true)
|-- source: string (nullable = true)
|-- text: string (nullable = true)
|-- timestamp_ms: string (nullable = true)
|-- truncated: boolean (nullable = true)
|-- tweet_text: string (nullable = true)
|-- user: struct (nullable = true)
|     |-- contributors_enabled: boolean (nullable = true)
|     |-- created_at: string (nullable = true)
|     |-- default_profile: boolean (nullable = true)
|     |-- default_profile_image: boolean (nullable = true)
|     |-- description: string (nullable = true)
|     |-- favourites_count: long (nullable = true)
|     |-- followers_count: long (nullable = true)
|     |-- friends_count: long (nullable = true)
|     |-- geo_enabled: boolean (nullable = true)
|     |-- id: long (nullable = true)
|     |-- id_str: string (nullable = true)
|     |-- is_translator: boolean (nullable = true)
|     |-- listed_count: long (nullable = true)
|     |-- location: string (nullable = true)
|     |-- name: string (nullable = true)
|     |-- profile_background_color: string (nullable = true)
|     |-- profile_background_image_url: string (nullable = true)
|     |-- profile_background_image_url_https: string (nullable = true)
|     |-- profile_background_tile: boolean (nullable = true)
|     |-- profile_banner_url: string (nullable = true)
|     |-- profile_image_url: string (nullable = true)
|     |-- profile_image_url_https: string (nullable = true)
|     |-- profile_link_color: string (nullable = true)
|     |-- profile_sidebar_border_color: string (nullable = true)
|     |-- profile_sidebar_fill_color: string (nullable = true)
|     |-- profile_text_color: string (nullable = true)
|     |-- profile_use_background_image: boolean (nullable = true)
|     |-- protected: boolean (nullable = true)
|     |-- screen_name: string (nullable = true)
|     |-- statuses_count: long (nullable = true)
|     |-- translator_type: string (nullable = true)
|     |-- url: string (nullable = true)
|     |-- verified: boolean (nullable = true)
|     |-- withheld_in_countries: array (nullable = true)
|     |     |-- element: string (containsNull = true)
|-- withheld_copyright: boolean (nullable = true)
|-- withheld_in_countries: array (nullable = true)
|     |-- element: string (containsNull = true)
```

In [10]:
```
#Selecting only the tweets that are in english
twitter_df=twitter_df.filter(twitter_df.lang=='en')
```

In [11]:
```
# Filtering tweets talking about primary, secondary and higher education
tweets_key=twitter_df.filter(lower(col('text')).contains('primary education')
                                  lower(col('text')).contains('secondary educat
                                 lower(col('text')).contains('primary school')|
                                 lower(col('text')).contains('elementary educat
                                 lower(col('text')).contains('normal school')|\
                                 lower(col('text')).contains('elementary school
```

```
                                lower(col('text')).contains('compulsory educat
                                lower(col('text')).contains('school')|\
                                lower(col('text')).contains('middle school')|\
                                lower(col('text')).contains('junior school')|\
                                lower(col('text')).contains('public school')|\
                                lower(col('text')).contains('education')|\
                                lower(col('text')).contains('educate')|\
                                lower(col('text')).contains('teacher')|\
                                lower(col('text')).contains('student')|\
                                lower(col('text')).contains('schoolmate')|\
                                lower(col('text')).contains('university')|\
                                lower(col('text')).contains('college')|\
                                lower(col('text')).contains('interschool')|\
                                lower(col('text')).contains('textbook')|\
                                lower(col('text')).contains('higher education'
                                lower(col('text')).contains('high school')|\
                                lower(col('text')).contains('students')|\
                                lower(col('text')).contains('schools')|\
                                lower(col('text')).contains('curriculum')|\
                                lower(col('text')).contains('undergraduate')|\
                                lower(col('text')).contains('grades')|\
                                lower(col('text')).contains('cgpa'))
```

In [82]:
```
tweets_key.count()
```

```
[Stage 118:===>       (1556 + 7) / 5401][Stage 119:>           (152 + 7) / 540
1]
```

In [54]:
```
null_count=tweets_df_key.select([count(when(col(c).isNull(), c)).alias(c) for
```

In [55]:
```
null_count=null_count.toPandas()
```

In [ ]:
```
no_rows=75674947
perc70_rows=0.7*no_rows

print("Total number of rows: ",no_rows)
print("70% of the total rows: ",perc70_rows)
```

```
[Stage 118:(2262 + 5) / 5401][Stage 119:(1205 + 5) / 5401][Stage 120:(577 + 4)
/ 5401]
```

In [10]:
```
perc70_col=[]
for col in null_count.columns:
    #print(col,null_count.loc[0,col])
    if null_count.loc[0,col]>perc70_rows:
        #print(col,null_count.loc[0,col])
        perc70_col.append(col)
perc70_col
```

Out[10]:
```
['coordinates',
 'display_text_range',
 'extended_entities',
 'extended_tweet',
 'geo',
 'in_reply_to_screen_name',
 'in_reply_to_status_id',
 'in_reply_to_status_id_str',
 'in_reply_to_user_id',
 'in_reply_to_user_id_str',
 'place',
 'possibly_sensitive',
 'quoted_status',
 'quoted_status_id',
 'quoted_status_id_str',
```

```
 'quoted_status_permalink',
 'quoted_text',
 'withheld_copyright',
 'withheld_in_countries']
```

In [4]:
```
tweets_df_key = tweets_df_key.drop('coordinates',
 'display_text_range',
 'extended_entities',
 'extended_tweet',
 'geo',
 'in_reply_to_screen_name',
 'in_reply_to_status_id',
 'in_reply_to_status_id_str',
 'in_reply_to_user_id',
 'in_reply_to_user_id_str',
 'place',
 'possibly_sensitive',
 'quoted_status',
 'quoted_status_id',
 'quoted_status_id_str',
 'quoted_status_permalink',
 'quoted_text',
 'withheld_copyright',
 'withheld_in_countries')
```

It is observed that the followings features have huge amounts of null values

Creating tweets data metastore

In [12]:
```
tweets_key.write.mode("overwrite").saveAsTable("tweets_table")
```

```
ivysettings.xml file not found in HIVE_HOME or HIVE_CONF_DIR,/etc/hive/conf.di
st/ivysettings.xml will be used
22/12/09 02:55:43 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 7 for reason Cont
ainer marked as failed: container_1670550927691_0003_01_000007 on host: hub-ms
ca-bdp-dphub-students-snigdag0402-sw-lmp1.c.msca-bdp-students.internal. Exit s
tatus: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:43 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 4 for reason Cont
ainer marked as failed: container_1670550927691_0003_01_000004 on host: hub-ms
ca-bdp-dphub-students-snigdag0402-sw-lmp1.c.msca-bdp-students.internal. Exit s
tatus: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:43 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 7 on hub-msca-bdp-dphub-students-snigdag0402-sw-lmp1.c.msca-bdp-stude
nts.internal: Container marked as failed: container_1670550927691_0003_01_0000
07 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-lmp1.c.msca-bdp-student
s.internal. Exit status: -100. Diagnostics: Container released on a *lost* nod
e.
22/12/09 02:55:43 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 4 on hub-msca-bdp-dphub-students-snigdag0402-sw-lmp1.c.msca-bdp-stude
nts.internal: Container marked as failed: container_1670550927691_0003_01_0000
04 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-lmp1.c.msca-bdp-student
s.internal. Exit status: -100. Diagnostics: Container released on a *lost* nod
e.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 15 on hub-msca-bdp-dphub-students-snigdag0402-sw-vvsw.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
015 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-vvsw.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 15 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000015 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-vvsw.c.msca-bdp-students.internal. Exit
```

```
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 10 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000010 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-vvsw.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 23 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000023 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-q7vs.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 18 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000018 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-q7vs.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 10 on hub-msca-bdp-dphub-students-snigdag0402-sw-vvsw.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
010 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-vvsw.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 23 on hub-msca-bdp-dphub-students-snigdag0402-sw-q7vs.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
023 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-q7vs.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 18 on hub-msca-bdp-dphub-students-snigdag0402-sw-q7vs.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
018 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-q7vs.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 13 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000013 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 20 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000020 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 14 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000014 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-l2kf.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 21 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000021 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-l2kf.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 13 on hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
013 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 20 on hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
020 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 14 on hub-msca-bdp-dphub-students-snigdag0402-sw-l2kf.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
```

```
014 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-l2kf.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 21 on hub-msca-bdp-dphub-students-snigdag0402-sw-l2kf.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
021 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-l2kf.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 02:55:54 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670550927691_0003_01_000012 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-qnxw.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-09 02:55:54.336]Container killed on request. Exit co
de is 143
[2022-12-09 02:55:54.338]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.375]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 12 for reason Con
tainer from a bad node: container_1670550927691_0003_01_000012 on host: hub-ms
ca-bdp-dphub-students-snigdag0402-sw-qnxw.c.msca-bdp-students.internal. Exit s
tatus: 143. Diagnostics: [2022-12-09 02:55:54.336]Container killed on request.
Exit code is 143
[2022-12-09 02:55:54.338]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.375]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670550927691_0003_01_000017 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-qnxw.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-09 02:55:54.338]Container killed on request. Exit co
de is 143
[2022-12-09 02:55:54.377]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.382]Killed by external signal
.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 12 on hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.c.msca-bdp-stud
ents.internal: Container from a bad node: container_1670550927691_0003_01_0000
12 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.c.msca-bdp-student
s.internal. Exit status: 143. Diagnostics: [2022-12-09 02:55:54.336]Container
killed on request. Exit code is 143
[2022-12-09 02:55:54.338]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.375]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 17 for reason Con
tainer from a bad node: container_1670550927691_0003_01_000017 on host: hub-ms
ca-bdp-dphub-students-snigdag0402-sw-qnxw.c.msca-bdp-students.internal. Exit s
tatus: 143. Diagnostics: [2022-12-09 02:55:54.338]Container killed on request.
Exit code is 143
[2022-12-09 02:55:54.377]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.382]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 30
9.0 in stage 2.0 (TID 15714) (hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.
c.msca-bdp-students.internal executor 12): ExecutorLostFailure (executor 12 ex
ited caused by one of the running tasks) Reason: Container from a bad node: co
ntainer_1670550927691_0003_01_000012 on host: hub-msca-bdp-dphub-students-snig
dag0402-sw-qnxw.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 02:55:54.336]Container killed on request. Exit code is 143
[2022-12-09 02:55:54.338]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.375]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.deploy.yarn.YarnAllocator: Cannot find
executorId for container: container_1670550927691_0003_01_000025
22/12/09 02:55:54 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 29
9.0 in stage 2.0 (TID 15704) (hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.
c.msca-bdp-students.internal executor 12): ExecutorLostFailure (executor 12 ex
ited caused by one of the running tasks) Reason: Container from a bad node: co
ntainer_1670550927691_0003_01_000012 on host: hub-msca-bdp-dphub-students-snig
```

```
dag0402-sw-qnxw.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 02:55:54.336]Container killed on request. Exit code is 143
[2022-12-09 02:55:54.338]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.375]Killed by external signal
.
22/12/09 02:55:54 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 17 on hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.c.msca-bdp-stud
ents.internal: Container from a bad node: container_1670550927691_0003_01_0000
17 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.c.msca-bdp-student
s.internal. Exit status: 143. Diagnostics: [2022-12-09 02:55:54.338]Container
killed on request. Exit code is 143
[2022-12-09 02:55:54.377]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.382]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 30
1.0 in stage 2.0 (TID 15706) (hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.
c.msca-bdp-students.internal executor 17): ExecutorLostFailure (executor 17 ex
ited caused by one of the running tasks) Reason: Container from a bad node: co
ntainer_1670550927691_0003_01_000017 on host: hub-msca-bdp-dphub-students-snig
dag0402-sw-qnxw.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 02:55:54.338]Container killed on request. Exit code is 143
[2022-12-09 02:55:54.377]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.382]Killed by external signal
.
22/12/09 02:55:54 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 32
9.1 in stage 2.0 (TID 15743) (hub-msca-bdp-dphub-students-snigdag0402-sw-qnxw.
c.msca-bdp-students.internal executor 17): ExecutorLostFailure (executor 17 ex
ited caused by one of the running tasks) Reason: Container from a bad node: co
ntainer_1670550927691_0003_01_000017 on host: hub-msca-bdp-dphub-students-snig
dag0402-sw-qnxw.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 02:55:54.338]Container killed on request. Exit code is 143
[2022-12-09 02:55:54.377]Container exited with a non-zero exit code 143.
[2022-12-09 02:55:54.382]Killed by external signal
.
22/12/09 03:13:19 WARN org.apache.hadoop.hive.ql.session.SessionState: METASTO
RE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is s
et to instance of HiveAuthorizerFactory.
```

In [57]:
```
#Total number of users
spark.sql("select count(distinct user.id) from tweets_table where").toPandas(
```

Out[57]:

	count(DISTINCT user.id AS)
0	16254090

In [56]:
```
#Total number of retweets
spark.sql("select count(*) from tweets_table where retweeted_status.retweet_c
```

Out[56]:

count(1)
48348031

In [ ]:

Question1

Identify the most prolific / influential Twitterers

- By message volume (original content)

- By message retweet (how often their messages are being retweeted). Please note: there are several variables that has “retweets count” in their name and you need to select correct one based on the EDA.
- Who are these Twitterers (government entities / universities / schools / nonprofit organizations / news outlets / social media influencers / someone else)?
- Visualize the distribution of tweet / retweet volume by Twitterers and types of organizations

Q1 - Part 1

In [88]:
```
msg_vol=spark.sql("select user.id_str, user.screen_name, sum(char_length(text
msg_vol
```

Out[88]:

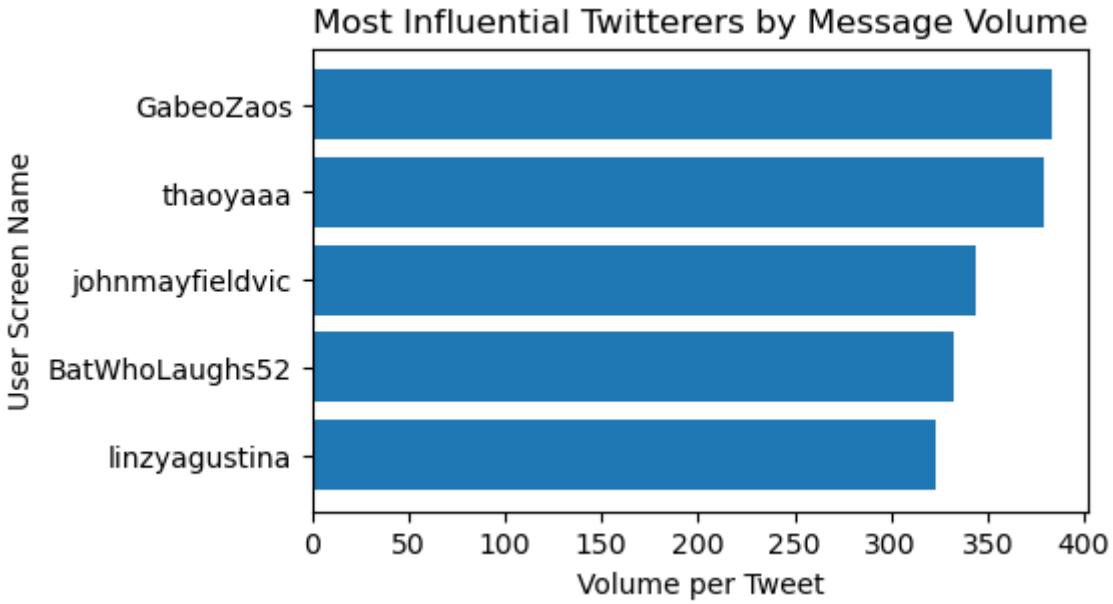
	id_str	screen_name	Volume
0	1458935628264095745	GabeoZaos	383.0
1	1273615219168653312	thaoyaaa	379.0
2	1547675376523939842	johnmayfieldvic	344.0
3	1155297287465328641	BatWhoLaughs52	332.0
4	1718044848	linzyagustina	323.0

In [ ]:
```
msg_vol=spark.sql("select user.id_str, user.screen_name, sum(char_length(text
msg_vol
```

```
[Stage 191:=================================================>  (926 + 41) / 96
7]
```

In [90]:
```
import pandas as pd
from matplotlib import pyplot as plt
```

In [133…
```
msg_vol= msg_vol.sort_values('Volume')

plt.figure(figsize=(5,3))
# bar plot with matplotlib
plt.barh('screen_name', 'Volume',data=msg_vol)
plt.xlabel("Volume per Tweet")
plt.ylabel("User Screen Name")
plt.title("Most Influential Twitterers by Message Volume", size=12)

plt.show()
```



Q1 - Part 2

In [107…
```
spark.sql("select count(distinct retweet_count) from tweets_table")
```

Out[107…

count(DISTINCT retweet_count)
1

In [17]:
```
spark.sql("select distinct retweet_count from tweets_table")
```

Out[17]:

retweet_count
0

In [108…
```
spark.sql("select count(distinct retweeted_status.retweet_count) from tweets_
```

Out[108…

count(DISTINCT retweeted_status.retweet_count AS $retweet_count$)
121878

In [109…
```
spark.sql("select count(distinct retweeted_status.quoted_status.retweet_count
```

Out[109…

count(DISTINCT retweeted_status.quoted_status.retweet_count AS $retweet_count$)
52753

In [21]:
```
spark.sql("select distinct retweeted_status.quoted_status.retweet_count from
```

Out[21]:

retweet_count
null

In [117…
```
retweet=spark.sql("select user.id_str, user.screen_name, sum(retweeted_status
          from tweets_table \
```

```
          group by user.id_str, user.screen_name \
          order by Total_Retweet_Count DESC \
          limit 5").toPandas()
retweet
```

```
22/12/08 12:14:55 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670471672595_0001_01_000125 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-08 12:14:55.352]Container killed on request. Exit co
de is 143
[2022-12-08 12:14:55.352]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.376]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670471672595_0001_01_000127 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-08 12:14:55.353]Container killed on request. Exit co
de is 143
[2022-12-08 12:14:55.376]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.377]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 123 for reason Co
ntainer from a bad node: container_1670471672595_0001_01_000125 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit
status: 143. Diagnostics: [2022-12-08 12:14:55.352]Container killed on reques
t. Exit code is 143
[2022-12-08 12:14:55.352]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.376]Killed by external signal
.
22/12/08 12:14:55 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 123 on hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-stu
dents.internal: Container from a bad node: container_1670471672595_0001_01_000
125 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-studen
ts.internal. Exit status: 143. Diagnostics: [2022-12-08 12:14:55.352]Container
killed on request. Exit code is 143
[2022-12-08 12:14:55.352]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.376]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 67
9.0 in stage 182.0 (TID 100560) (hub-msca-bdp-dphub-students-snigdag0402-sw-zw
k1.c.msca-bdp-students.internal executor 123): ExecutorLostFailure (executor 1
23 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000125 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:14:55.352]Container killed on request. Exit code is 143
[2022-12-08 12:14:55.352]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.376]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 67
5.0 in stage 182.0 (TID 100556) (hub-msca-bdp-dphub-students-snigdag0402-sw-zw
k1.c.msca-bdp-students.internal executor 123): ExecutorLostFailure (executor 1
23 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000125 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:14:55.352]Container killed on request. Exit code is 143
[2022-12-08 12:14:55.352]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.376]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 125 for reason Co
ntainer from a bad node: container_1670471672595_0001_01_000127 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit
status: 143. Diagnostics: [2022-12-08 12:14:55.353]Container killed on reques
t. Exit code is 143
[2022-12-08 12:14:55.376]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.377]Killed by external signal
.
22/12/08 12:14:55 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
```

```
executor 125 on hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-stu
dents.internal: Container from a bad node: container_1670471672595_0001_01_000
127 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-studen
ts.internal. Exit status: 143. Diagnostics: [2022-12-08 12:14:55.353]Container
killed on request. Exit code is 143
[2022-12-08 12:14:55.376]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.377]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.deploy.yarn.YarnAllocator: Cannot find
executorId for container: container_1670471672595_0001_01_000130
22/12/08 12:14:55 WARN org.apache.spark.deploy.yarn.YarnAllocator: Cannot find
executorId for container: container_1670471672595_0001_01_000129
22/12/08 12:14:55 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 67
7.0 in stage 182.0 (TID 100558) (hub-msca-bdp-dphub-students-snigdag0402-sw-zw
k1.c.msca-bdp-students.internal executor 125): ExecutorLostFailure (executor 1
25 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000127 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:14:55.353]Container killed on request. Exit code is 143
[2022-12-08 12:14:55.376]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.377]Killed by external signal
.
22/12/08 12:14:55 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 67
3.0 in stage 182.0 (TID 100554) (hub-msca-bdp-dphub-students-snigdag0402-sw-zw
k1.c.msca-bdp-students.internal executor 125): ExecutorLostFailure (executor 1
25 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000127 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zwk1.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:14:55.353]Container killed on request. Exit code is 143
[2022-12-08 12:14:55.376]Container exited with a non-zero exit code 143.
[2022-12-08 12:14:55.377]Killed by external signal
.
22/12/08 12:17:24 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 4.
0 in stage 184.0 (TID 100856) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.m
sca-bdp-students.internal executor 120): FetchFailed(BlockManagerId(125, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 7337,
None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=16, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839000,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
```

```
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
```

```
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839000,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
```

```
	at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
	at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
	at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
	at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
	at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
	at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
	at java.lang.Thread.run(Thread.java:750)

	at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
	at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
	at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
	at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
	at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
	at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
	at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
	at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
	at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
	at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
	at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
```

```
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:24 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 1.
0 in stage 184.0 (TID 100853) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.m
sca-bdp-students.internal executor 120): FetchFailed(BlockManagerId(123, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 7337,
None), shuffleId=52, mapIndex=40, mapId=99921, reduceId=4, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839001,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
```

```
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
```

```
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839001,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
```

```
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
```

```
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:24 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 2.
0 in stage 184.0 (TID 100854) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.m
sca-bdp-students.internal executor 126): FetchFailed(BlockManagerId(125, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 7337,
None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=8, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839006,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
```

```
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
```

```
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839006,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
```

```
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
```

```
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:24 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 5.
0 in stage 184.0 (TID 100857) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.m
sca-bdp-students.internal executor 126): FetchFailed(BlockManagerId(125, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 7337,
None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=20, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839007,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
```

```
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
```

```
	at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
	at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
	at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
	at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
	at org.apache.spark.scheduler.Task.run(Task.scala:131)
	at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
	at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
	at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
	at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
	at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
	at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839007,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=125)
	at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
	at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
	at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
	at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
	at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
	at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
	at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
	at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
	at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
	at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
	at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
	at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
	at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
```

```
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
```

```
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:24 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 3.
0 in stage 184.0 (TID 100855) (hub-msca-bdp-dphub-students-snigdag0402-w-0.c.m
sca-bdp-students.internal executor 124): FetchFailed(BlockManagerId(125, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 7337,
None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=12, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839008,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
```

```
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
```

```
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Executor.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:508)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure while fetching StreamChunkId[streamId=1315162839008,chunkIndex=0]: java.lang.RuntimeException: Executor is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getContinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManagedBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManagedBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(OneForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFetchRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(TransportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(TransportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
```

```
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
```

```
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:24 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 0.
0 in stage 184.0 (TID 100852) (hub-msca-bdp-dphub-students-snigdag0402-w-0.c.m
sca-bdp-students.internal executor 124): FetchFailed(BlockManagerId(123, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 7337,
None), shuffleId=52, mapIndex=40, mapId=99921, reduceId=0, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839009,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
```

```
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
	at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
	at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
	at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
	at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
	at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
	at java.lang.Thread.run(Thread.java:750)

	at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
	at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
	at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
	at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
	at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
	at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
	at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
	at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
	at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
	at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
	at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
	at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
	at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
	at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
	at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
	at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
	at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
	at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
	at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
	at org.apache.spark.scheduler.Task.run(Task.scala:131)
```

```
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Executor.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:508)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure while fetching StreamChunkId[streamId=1315162839009,chunkIndex=0]: java.lang.RuntimeException: Executor is not registered (appId=application_1670471672595_0001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getContinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManagedBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManagedBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(OneForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFetchRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(TransportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(TransportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
```

```
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
```

```
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:26 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 1
0.0 in stage 184.0 (TID 100862) (hub-msca-bdp-dphub-students-snigdag0402-w-1.
c.msca-bdp-students.internal executor 120): FetchFailed(BlockManagerId(123, hu
b-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal, 73
37, None), shuffleId=52, mapIndex=40, mapId=99921, reduceId=40, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839004,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
```

```
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
```

```
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839004,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
```

```
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
```

```
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:28 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 7.
0 in stage 184.0 (TID 100859) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt2
m.c.msca-bdp-students.internal executor 122): FetchFailed(BlockManagerId(125,
 hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal,
7337, None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=28, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839014,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
```

```
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
	at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
	at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
	at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
	at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
	at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
	at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
	at java.lang.Thread.run(Thread.java:750)

	at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
	at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
	at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
	at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
	at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
	at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
	at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
	at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
	at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
	at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
	at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
	at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
	at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
	at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
	at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
	at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
	at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
	at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
	at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
	at org.apache.spark.scheduler.Task.run(Task.scala:131)
	at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
	at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
	at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
	at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
```

```
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839014,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
```

```
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
```

```
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:28 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 9.
0 in stage 184.0 (TID 100861) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt2
m.c.msca-bdp-students.internal executor 122): FetchFailed(BlockManagerId(125,
 hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal,
7337, None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=36, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839015,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
```

```
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
```

```
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839015,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
```

```
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
```

```
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
22/12/08 12:17:28 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 8.
0 in stage 184.0 (TID 100860) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt2
m.c.msca-bdp-students.internal executor 121): FetchFailed(BlockManagerId(125,
 hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal,
7337, None), shuffleId=52, mapIndex=34, mapId=99915, reduceId=32, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839018,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
```

```
	at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
	at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
	at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
	at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
	at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
	at java.lang.Thread.run(Thread.java:750)

	at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
	at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
	at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
	at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
	at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
	at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
	at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
	at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
	at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
	at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
	at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
	at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
	at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
	at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
	at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
	at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
	at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
	at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
	at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
	at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
	at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
	at org.apache.spark.scheduler.Task.run(Task.scala:131)
	at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
	at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
	at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
	at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
	at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
	at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839018,chunkIndex=0]: java.lang.R
```

```
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=125)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
```

```
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more
```

```
)
22/12/08 12:17:28 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 6.
0 in stage 184.0 (TID 100858) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt2
m.c.msca-bdp-students.internal executor 121): FetchFailed(BlockManagerId(123,
 hub-msca-bdp-dphub-students-snigdag0402-sw-zwk1.c.msca-bdp-students.internal,
7337, None), shuffleId=52, mapIndex=40, mapId=99921, reduceId=24, message=
org.apache.spark.shuffle.FetchFailedException: Failure while fetching StreamCh
unkId[streamId=1315162839019,chunkIndex=0]: java.lang.RuntimeException: Execut
or is not registered (appId=application_1670471672595_0001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
```

```
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:777)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithKeys_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at scala.collection.convert.Wrappers$IteratorWrapper.hasNext(Wrappers.
scala:32)
        at org.sparkproject.guava.collect.Ordering.leastOf(Ordering.java:763)
        at org.apache.spark.util.collection.Utils$.takeOrdered(Utils.scala:37)
        at org.apache.spark.rdd.RDD.$anonfun$takeOrdered$2(RDD.scala:1518)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2(RDD.scala:863)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitions$2$adapted(RDD.scal
a:863)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: org.apache.spark.network.client.ChunkFetchFailureException: Failure
while fetching StreamChunkId[streamId=1315162839019,chunkIndex=0]: java.lang.R
untimeException: Executor is not registered (appId=application_1670471672595_0
001, execId=123)
        at org.apache.spark.network.shuffle.ExternalShuffleBlockResolver.getCo
```

```
ntinuousBlocksData(ExternalShuffleBlockResolver.java:188)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:489)
        at org.apache.spark.network.shuffle.ExternalBlockHandler$ShuffleManage
dBufferIterator.next(ExternalBlockHandler.java:445)
        at org.apache.spark.network.server.OneForOneStreamManager.getChunk(One
ForOneStreamManager.java:87)
        at org.apache.spark.network.server.ChunkFetchRequestHandler.processFet
chRequest(ChunkFetchRequestHandler.java:103)
        at org.apache.spark.network.server.TransportRequestHandler.handle(Tran
sportRequestHandler.java:107)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:140)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at org.sparkproject.io.netty.channel.SimpleChannelInboundHandler.chann
elRead(SimpleChannelInboundHandler.java:99)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.timeout.IdleStateHandler.channelR
ead(IdleStateHandler.java:286)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.handler.codec.MessageToMessageDecoder.cha
nnelRead(MessageToMessageDecoder.java:103)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.fir
eChannelRead(AbstractChannelHandlerContext.java:357)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline$HeadContex
t.channelRead(DefaultChannelPipeline.java:1410)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:379)
        at org.sparkproject.io.netty.channel.AbstractChannelHandlerContext.inv
okeChannelRead(AbstractChannelHandlerContext.java:365)
        at org.sparkproject.io.netty.channel.DefaultChannelPipeline.fireChanne
lRead(DefaultChannelPipeline.java:919)
        at org.sparkproject.io.netty.channel.nio.AbstractNioByteChannel$NioByt
eUnsafe.read(AbstractNioByteChannel.java:163)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
ey(NioEventLoop.java:714)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eysOptimized(NioEventLoop.java:650)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.processSelectedK
eys(NioEventLoop.java:576)
        at org.sparkproject.io.netty.channel.nio.NioEventLoop.run(NioEventLoo
p.java:493)
        at org.sparkproject.io.netty.util.concurrent.SingleThreadEventExecutor
$4.run(SingleThreadEventExecutor.java:989)
        at org.sparkproject.io.netty.util.internal.ThreadExecutorMap$2.run(Thr
eadExecutorMap.java:74)
```

```
        at org.sparkproject.io.netty.util.concurrent.FastThreadLocalRunnable.r
un(FastThreadLocalRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

        at org.apache.spark.network.client.TransportResponseHandler.handle(Tra
nsportResponseHandler.java:182)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:142)
        at org.apache.spark.network.server.TransportChannelHandler.channelRead
0(TransportChannelHandler.java:53)
        at io.netty.channel.SimpleChannelInboundHandler.channelRead(SimpleChan
nelInboundHandler.java:99)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.timeout.IdleStateHandler.channelRead(IdleStateHand
ler.java:286)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.handler.codec.MessageToMessageDecoder.channelRead(MessageT
oMessageDecoder.java:103)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at org.apache.spark.network.util.TransportFrameDecoder.channelRead(Tra
nsportFrameDecoder.java:102)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(Abst
ractChannelHandlerContext.java:357)
        at io.netty.channel.DefaultChannelPipeline$HeadContext.channelRead(Def
aultChannelPipeline.java:1410)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:379)
        at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(Ab
stractChannelHandlerContext.java:365)
        at io.netty.channel.DefaultChannelPipeline.fireChannelRead(DefaultChan
nelPipeline.java:919)
        at io.netty.channel.nio.AbstractNioByteChannel$NioByteUnsafe.read(Abst
ractNioByteChannel.java:163)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:714)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more

)
```

Out[117…

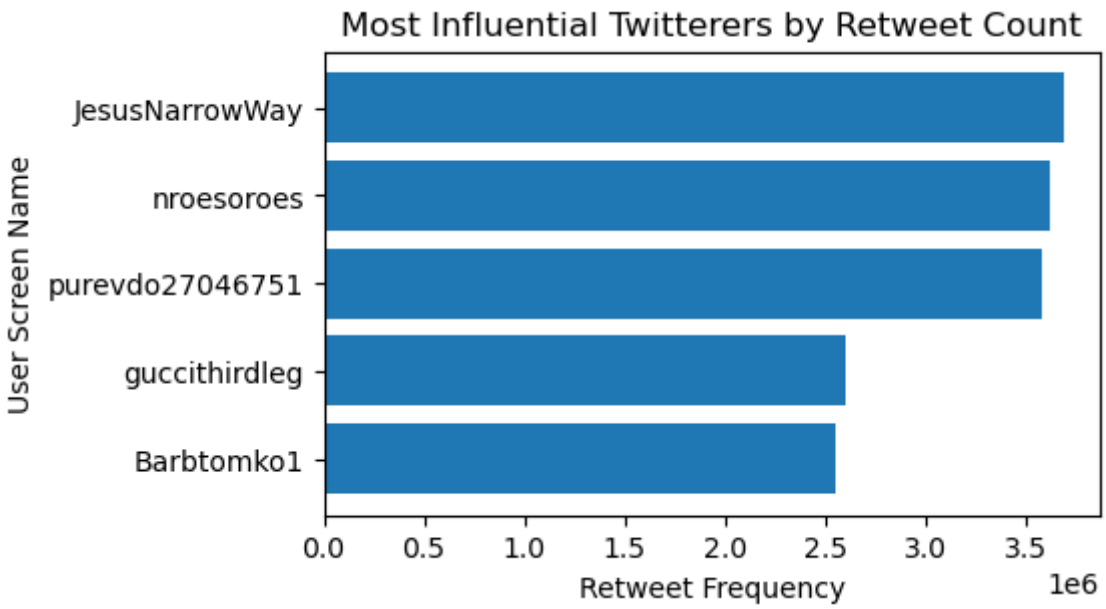
	id_str	screen_name	Total_Retweet_Count
0	362084030	JesusNarrowWay	3688172
1	1516373316033863690	nroesoroes	3623418
2	1552239312195817474	purevdo27046751	3578311
3	1501043135639695367	guccithirdleg	2602455
4	804430039	Barbtomko1	2552806

In [118…
```
retweet
```

Out[118…

	id_str	screen_name	Total_Retweet_Count
0	362084030	JesusNarrowWay	3688172
1	1516373316033863690	nroesoroes	3623418
2	1552239312195817474	purevdo27046751	3578311
3	1501043135639695367	guccithirdleg	2602455
4	804430039	Barbtomko1	2552806

In [132…
```
retweet= retweet.sort_values('Total_Retweet_Count')

plt.figure(figsize=(5,3))
# bar plot with matplotlib
plt.barh('screen_name', 'Total_Retweet_Count',data=retweet)
plt.xlabel("Retweet Frequency")
plt.ylabel("User Screen Name")
plt.title("Most Influential Twitterers by Retweet Count", size=12)
plt.show()
```



Q1- Part 3

In [ ]:
```
government entities / universities / schools / nonprofit organizations / news
```

In [ ]:
```

```

In [136…
```
df1=spark.sql("select 'social media influencer' as background, count(*) as co
```

In [137…

```
df2=spark.sql("select 'government entities' as background, count(*) as count
```

```
22/12/08 12:56:23 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670471672595_0001_01_000164 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-08 12:56:23.241]Container killed on request. Exit co
de is 143
[2022-12-08 12:56:23.242]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670471672595_0001_01_000165 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-08 12:56:23.241]Container killed on request. Exit co
de is 143
[2022-12-08 12:56:23.241]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 158 for reason Co
ntainer from a bad node: container_1670471672595_0001_01_000164 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit
status: 143. Diagnostics: [2022-12-08 12:56:23.241]Container killed on reques
t. Exit code is 143
[2022-12-08 12:56:23.242]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.deploy.yarn.YarnAllocator: Cannot find
executorId for container: container_1670471672595_0001_01_000168
22/12/08 12:56:23 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 159 for reason Co
ntainer from a bad node: container_1670471672595_0001_01_000165 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit
status: 143. Diagnostics: [2022-12-08 12:56:23.241]Container killed on reques
t. Exit code is 143
[2022-12-08 12:56:23.241]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 158 on hub-msca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-stu
dents.internal: Container from a bad node: container_1670471672595_0001_01_000
164 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-studen
ts.internal. Exit status: 143. Diagnostics: [2022-12-08 12:56:23.241]Container
killed on request. Exit code is 143
[2022-12-08 12:56:23.242]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.deploy.yarn.YarnAllocator: Cannot find
executorId for container: container_1670471672595_0001_01_000169
22/12/08 12:56:23 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 8
2.0 in stage 197.0 (TID 104960) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt
2m.c.msca-bdp-students.internal executor 158): ExecutorLostFailure (executor 1
58 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000164 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:56:23.241]Container killed on request. Exit code is 143
[2022-12-08 12:56:23.242]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 8
7.0 in stage 197.0 (TID 104965) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt
2m.c.msca-bdp-students.internal executor 158): ExecutorLostFailure (executor 1
58 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000164 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
```

```
cs: [2022-12-08 12:56:23.241]Container killed on request. Exit code is 143
[2022-12-08 12:56:23.242]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 159 on hub-msca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-stu
dents.internal: Container from a bad node: container_1670471672595_0001_01_000
165 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-studen
ts.internal. Exit status: 143. Diagnostics: [2022-12-08 12:56:23.241]Container
killed on request. Exit code is 143
[2022-12-08 12:56:23.241]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 8
6.0 in stage 197.0 (TID 104964) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt
2m.c.msca-bdp-students.internal executor 159): ExecutorLostFailure (executor 1
59 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000165 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:56:23.241]Container killed on request. Exit code is 143
[2022-12-08 12:56:23.241]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 9
1.0 in stage 197.0 (TID 104969) (hub-msca-bdp-dphub-students-snigdag0402-sw-zt
2m.c.msca-bdp-students.internal executor 159): ExecutorLostFailure (executor 1
59 exited caused by one of the running tasks) Reason: Container from a bad nod
e: container_1670471672595_0001_01_000165 on host: hub-msca-bdp-dphub-students
-snigdag0402-sw-zt2m.c.msca-bdp-students.internal. Exit status: 143. Diagnosti
cs: [2022-12-08 12:56:23.241]Container killed on request. Exit code is 143
[2022-12-08 12:56:23.241]Container exited with a non-zero exit code 143.
[2022-12-08 12:56:23.242]Killed by external signal
.
22/12/08 12:56:23 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Igno
ring update with state FINISHED for TID 104964 because its task set is gone (t
his is likely the result of receiving duplicate task finished status updates)
or its executor has been marked as failed.
22/12/08 12:56:23 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnDriverEndpoint: Ignored task status update (104964 state FINISHED) from u
nknown executor with ID 159
22/12/08 12:58:18 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 0.
0 in stage 199.0 (TID 105849) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.m
sca-bdp-students.internal executor 156): FetchFailed(BlockManagerId(158, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal, 7337,
None), shuffleId=57, mapIndex=38, mapId=104916, reduceId=0, message=
org.apache.spark.shuffle.FetchFailedException: Failed to connect to hub-msca-b
dp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal/10.128.0.5
5:7337
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:775)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithoutKey_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
```

```
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at org.apache.spark.sql.execution.SparkPlan.$anonfun$getByteArrayRdd$1
(SparkPlan.scala:345)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitionsInternal$2(RDD.scal
a:898)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitionsInternal$2$adapted(R
DD.scala:898)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: java.io.IOException: Failed to connect to hub-msca-bdp-dphub-studen
ts-snigdag0402-sw-zt2m.c.msca-bdp-students.internal/10.128.0.55:7337
        at org.apache.spark.network.client.TransportClientFactory.createClient
(TransportClientFactory.java:287)
        at org.apache.spark.network.client.TransportClientFactory.createClient
(TransportClientFactory.java:218)
        at org.apache.spark.network.shuffle.ExternalBlockStoreClient.lambda$fe
tchBlocks$0(ExternalBlockStoreClient.java:105)
        at org.apache.spark.network.shuffle.RetryingBlockFetcher.fetchAllOutst
anding(RetryingBlockFetcher.java:153)
        at org.apache.spark.network.shuffle.RetryingBlockFetcher.lambda$initia
teRetry$0(RetryingBlockFetcher.java:181)
        at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:
511)
        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more
Caused by: io.netty.channel.AbstractChannel$AnnotatedConnectException: Connect
ion refused: hub-msca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-studen
ts.internal/10.128.0.55:7337
Caused by: java.net.ConnectException: Connection refused
        at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
        at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:7
16)
        at io.netty.channel.socket.nio.NioSocketChannel.doFinishConnect(NioSoc
ketChannel.java:330)
        at io.netty.channel.nio.AbstractNioChannel$AbstractNioUnsafe.finishCon
nect(AbstractNioChannel.java:334)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:702)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
```

```
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

)
22/12/08 12:58:33 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 0.
0 in stage 199.1 (TID 105852) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.m
sca-bdp-students.internal executor 156): FetchFailed(BlockManagerId(159, hub-m
sca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal, 7337,
None), shuffleId=57, mapIndex=39, mapId=104917, reduceId=0, message=
org.apache.spark.shuffle.FetchFailedException: Failed to connect to hub-msca-b
dp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-students.internal/10.128.0.5
5:7337
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.throwFetchFail
edException(ShuffleBlockFetcherIterator.scala:775)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:690)
        at org.apache.spark.storage.ShuffleBlockFetcherIterator.next(ShuffleBl
ockFetcherIterator.scala:70)
        at org.apache.spark.util.CompletionIterator.next(CompletionIterator.sc
ala:29)
        at scala.collection.Iterator$$anon$11.nextCur(Iterator.scala:486)
        at scala.collection.Iterator$$anon$11.hasNext(Iterator.scala:492)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.util.CompletionIterator.hasNext(CompletionIterato
r.scala:31)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterato
r.scala:37)
        at scala.collection.Iterator$$anon$10.hasNext(Iterator.scala:460)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.agg_doAggregateWithoutKey_0$(Unknown Source)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage2.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at org.apache.spark.sql.execution.SparkPlan.$anonfun$getByteArrayRdd$1
(SparkPlan.scala:345)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitionsInternal$2(RDD.scal
a:898)
        at org.apache.spark.rdd.RDD.$anonfun$mapPartitionsInternal$2$adapted(R
DD.scala:898)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: java.io.IOException: Failed to connect to hub-msca-bdp-dphub-studen
ts-snigdag0402-sw-zt2m.c.msca-bdp-students.internal/10.128.0.55:7337
        at org.apache.spark.network.client.TransportClientFactory.createClient
(TransportClientFactory.java:287)
        at org.apache.spark.network.client.TransportClientFactory.createClient
(TransportClientFactory.java:218)
        at org.apache.spark.network.shuffle.ExternalBlockStoreClient.lambda$fe
tchBlocks$0(ExternalBlockStoreClient.java:105)
        at org.apache.spark.network.shuffle.RetryingBlockFetcher.fetchAllOutst
anding(RetryingBlockFetcher.java:153)
```

```
        at org.apache.spark.network.shuffle.RetryingBlockFetcher.lambda$initia
teRetry$0(RetryingBlockFetcher.java:181)
        at java.util.concurrent.Executors$RunnableAdapter.call(Executors.java:
511)
        at java.util.concurrent.FutureTask.run(FutureTask.java:266)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        ... 1 more
Caused by: io.netty.channel.AbstractChannel$AnnotatedConnectException: Connect
ion refused: hub-msca-bdp-dphub-students-snigdag0402-sw-zt2m.c.msca-bdp-studen
ts.internal/10.128.0.55:7337
Caused by: java.net.ConnectException: Connection refused
        at sun.nio.ch.SocketChannelImpl.checkConnect(Native Method)
        at sun.nio.ch.SocketChannelImpl.finishConnect(SocketChannelImpl.java:7
16)
        at io.netty.channel.socket.nio.NioSocketChannel.doFinishConnect(NioSoc
ketChannel.java:330)
        at io.netty.channel.nio.AbstractNioChannel$AbstractNioUnsafe.finishCon
nect(AbstractNioChannel.java:334)
        at io.netty.channel.nio.NioEventLoop.processSelectedKey(NioEventLoop.j
ava:702)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeysOptimized(NioE
ventLoop.java:650)
        at io.netty.channel.nio.NioEventLoop.processSelectedKeys(NioEventLoop.
java:576)
        at io.netty.channel.nio.NioEventLoop.run(NioEventLoop.java:493)
        at io.netty.util.concurrent.SingleThreadEventExecutor$4.run(SingleThre
adEventExecutor.java:989)
        at io.netty.util.internal.ThreadExecutorMap$2.run(ThreadExecutorMap.ja
va:74)
        at io.netty.util.concurrent.FastThreadLocalRunnable.run(FastThreadLoca
lRunnable.java:30)
        at java.lang.Thread.run(Thread.java:750)

)
```

In [138…

```
df3=pd.concat([df1, df2])
```

Out[138…

	background	count
0	social media influencer	89703
0	government entities	248966

```
22/12/08 12:58:37 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670471672595_0001_01_000151 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-0wvj.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-08 12:58:37.144]Container killed on request. Exit co
de is 143
[2022-12-08 12:58:37.144]Container exited with a non-zero exit code 143.
[2022-12-08 12:58:37.145]Killed by external signal
.
22/12/08 12:58:37 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 145 for reason Co
ntainer from a bad node: container_1670471672595_0001_01_000151 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-0wvj.c.msca-bdp-students.internal. Exit
status: 143. Diagnostics: [2022-12-08 12:58:37.144]Container killed on reques
t. Exit code is 143
[2022-12-08 12:58:37.144]Container exited with a non-zero exit code 143.
[2022-12-08 12:58:37.145]Killed by external signal
.
22/12/08 12:58:37 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 145 on hub-msca-bdp-dphub-students-snigdag0402-sw-0wvj.c.msca-bdp-stu
```

```
dents.internal: Container from a bad node: container_1670471672595_0001_01_000
151 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-0wvj.c.msca-bdp-studen
ts.internal. Exit status: 143. Diagnostics: [2022-12-08 12:58:37.144]Container
killed on request. Exit code is 143
[2022-12-08 12:58:37.144]Container exited with a non-zero exit code 143.
[2022-12-08 12:58:37.145]Killed by external signal
.
```

In [139…
```
df3
```

Out[139…

	background	count
0	social media influencer	89703
0	government entities	248966

In [ ]:
```
df4=spark.sql("select 'universities' as background, count(*) as count from tw
```

In [ ]:
```
df5=spark.sql("select 'schools' as background, count(*) as count from tweets_
```

In [ ]:
```
df6=spark.sql("select 'nonprofit organizations' as background, count(*) as co
```

In [ ]:
```
df7=spark.sql("select 'news outlets' as background, count(*) as count from tw
or lower(user.description) like '%media%'").toPandas()
```

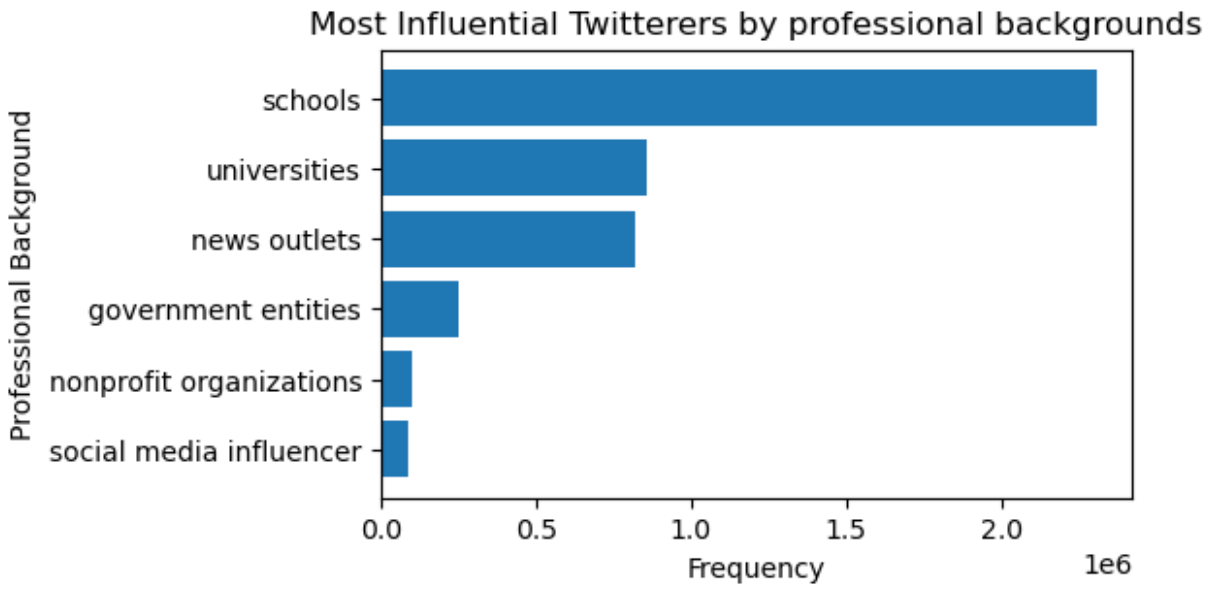
In [148…
```
prof_bg=pd.concat([df3, df4,df5, df6,df7])
```

In [152…
```
prof_bg.sort_values('count')
```

Out[152…

	background	count
0	social media influencer	89703
0	nonprofit organizations	100866
0	government entities	248966
0	news outlets	818864
0	universities	858565
0	schools	2308094

In [151…
```
prof_bg= prof_bg.sort_values('count')

plt.figure(figsize=(5,3))
# bar plot with matplotlib
plt.barh('background', 'count',data=prof_bg)
plt.xlabel("Frequency")
plt.ylabel("Professional Background")
plt.title("Most Influential Twitterers by professional backgrounds", size=12)
plt.show()
```



In [ ]:

Q1 - PART 4

In [155…

```
q1_p4=spark.sql("select case when lower(user.description) like '%news outlet%
          when lower(user.description) like '%non%profit%' then 'nonprofit or
          when lower(user.description) like '%school%' then 'schools' \
          when lower(user.description) like '%universit%' then 'universities'
          when lower(user.description) like '%government%' then 'government e
          when lower(user.description) like '%influencer%' then 'social media
          end as pf_bg, \
          sum(retweeted_status.retweet_count) as Total_Retweet_Count, \
          sum(char_length(text)) as Volume \
          from tweets_table \
          group by pf_bg").toPandas()
q1_p4
```

Out[155…

	pf_bg	Total_Retweet_Count	Volume
0	schools	503120263	294083309
1	social media influencer	49871848	7888800
2	None	120493716412	8759185513
3	universities	549272378	98080781
4	government entities	188454315	29777504
5	news outlets	646457982	103633968
6	nonprofit organizations	46703835	13001086

In [160…

```
q1_p4 = q1_p4.drop(labels=[2], axis=0)
q1_p4
```

Out[160…

	pf_bg	Total_Retweet_Count	Volume
1	social media influencer	49871848	7888800
6	nonprofit organizations	46703835	13001086

	pf_bg	Total_Retweet_Count	Volume
4	government entities	188454315	29777504
3	universities	549272378	98080781
5	news outlets	646457982	103633968
0	schools	503120263	294083309

In [168…

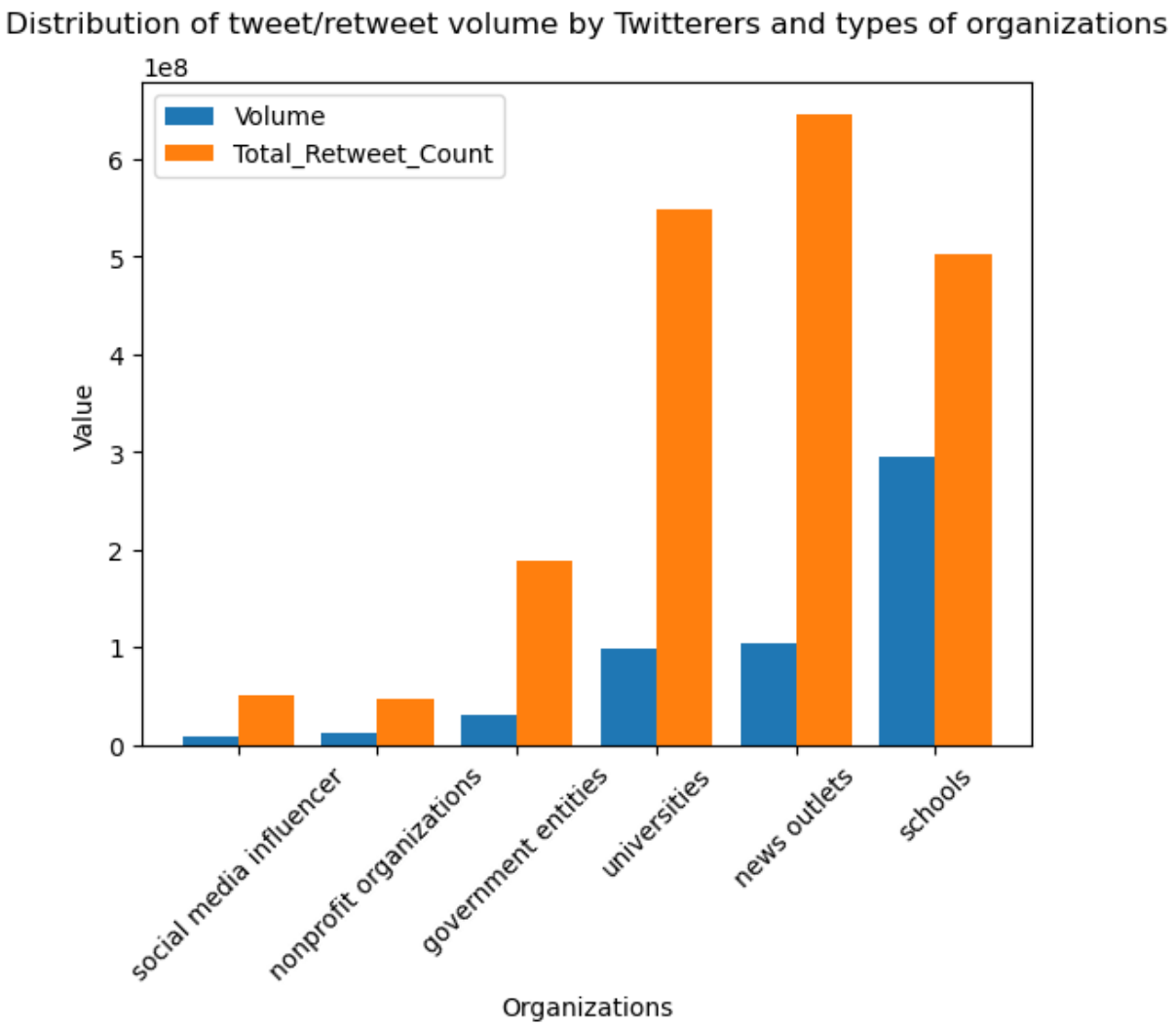
```
q1_p4=q1_p4.sort_values('Volume',ascending=False)
q1_p4
```

Out[168…

	pf_bg	Total_Retweet_Count	Volume
0	schools	503120263	294083309
5	news outlets	646457982	103633968
3	universities	549272378	98080781
4	government entities	188454315	29777504
6	nonprofit organizations	46703835	13001086
1	social media influencer	49871848	7888800

In [166…

```
import numpy as np
#import matplotlib.pyplot as plt

from matplotlib import rcParams
rcParams['axes.titlepad'] = 20

X = list(q1_p4["pf_bg"])
V = list(q1_p4["Volume"])
R = list(q1_p4["Total_Retweet_Count"])

X_axis = np.arange(len(X))

plt.bar(X_axis - 0.2, V, 0.4, label = 'Volume')
plt.bar(X_axis + 0.2, R, 0.4, label = 'Total_Retweet_Count')

plt.xticks(X_axis, X)
plt.xlabel("Organizations")
plt.xticks(rotation = 45)
plt.ylabel("Value")
plt.title("Distribution of tweet/retweet volume by Twitterers and types of or
plt.legend()
plt.show()
```



In [ ]:

Question 2

Where are these Twitterers (all of them, not just influencers) located?

- Do you see any relationship between the emergence of new issues in education and progression and locations of these Twitterers?
- Visualize the geographical distribution

In [178…
```
import pyspark

split_col = pyspark.sql.functions.split(tweets_key['user.location'], ',')
tweets_key = tweets_key.withColumn('user_place', split_col.getItem(0))
tweets_key = tweets_key.withColumn('user_country', split_col.getItem(1))
```

In [ ]:
```
tweets_key.write.mode("overwrite").saveAsTable("tweets_table2")
```

In [ ]:
```
spark.sql("select user.location,user_place,user_country from tweets_table2 li
```

Out[ ]:

	location	user_place	user_country

	location	user_place	user_country
0	None	None	None
1	Ann Arbor, MI	Ann Arbor	MI
2	None	None	None
3	Lagos, Nigeria	Lagos	Nigeria
4	Canajoharie, NY	Canajoharie	NY
5	None	None	None
6	North	North	None
7	Stamford, CT	Stamford	CT
8	Reedley, CA	Reedley	CA
9	Reykjavik, Iceland	Reykjavik	Iceland
10	None	None	None
11	Pobal O Caoimh, County Cork - Bridge River Valley - Vancouver	Pobal O Caoimh	County Cork - Bridge River Valley - Vancouver
12	None	None	None
13	None	None	None
14	Atlanta, GA	Atlanta	GA
15	El Paso, TX	El Paso	TX
16	NY	NY	None
17	California - Bay area	California - Bay area	None
18	None	None	None
19	Valley Stream, NY	Valley Stream	NY
20	Once Great State of Texas	Once Great State of Texas	None
21	Christchurch City, New Zealand	Christchurch City	New Zealand
22	Canada	Canada	None
23	Eugene	Eugene	None
24	In the path of the Artic Blast	In the path of the Artic Blast	None
25	Nigeria	Nigeria	None
26	Austin, Texas	Austin	Texas
27	Cape Town, South Africa	Cape Town	South Africa
28	manchester	manchester	None
29	New Mexico, USA	New Mexico	USA

In [184…
```
loc=spark.sql("select user_place, count(*) as count from tweets_table2 where
```

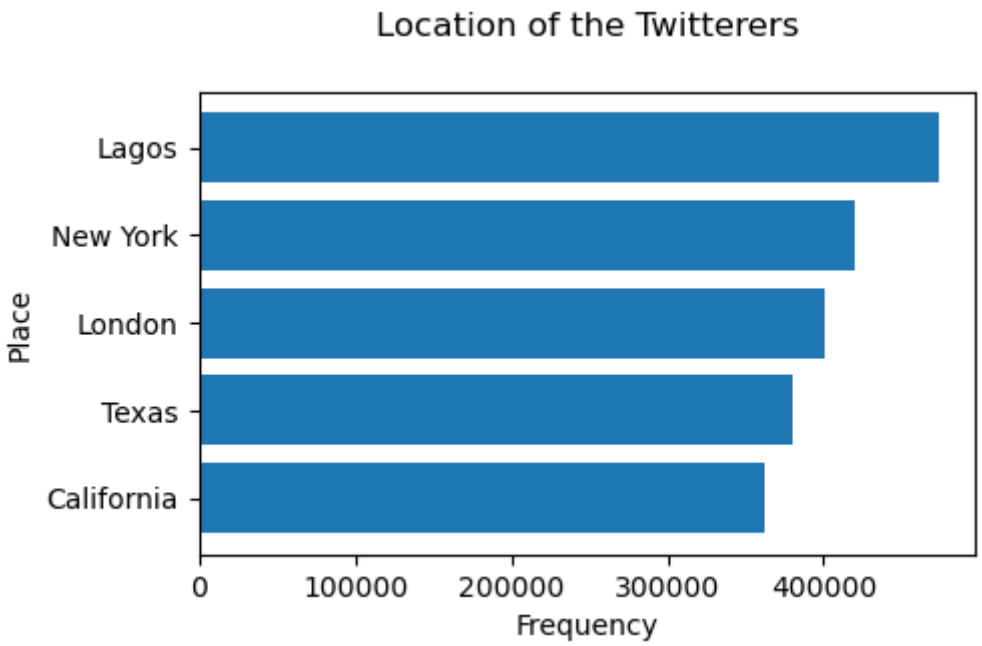
In [185…
```
loc
```

Out[185…

	user_place	count
0	Lagos	473387
1	New York	419528
2	London	400238
3	Texas	380219
4	California	361585

In [186…

```
loc= loc.sort_values('count')

plt.figure(figsize=(5,3))
# bar plot with matplotlib
plt.barh('user_place', 'count',data=loc)
plt.xlabel("Frequency")
plt.ylabel("Place")
plt.title("Location of the Twitterers", size=12)
plt.show()
```



In [196…

```
loc2=spark.sql("select user_place, count(*) as count \
from tweets_table2 \
where user_place is not null and text like '%book%ban%'\
group by user_place \
order by count desc limit 5").toPandas()
```

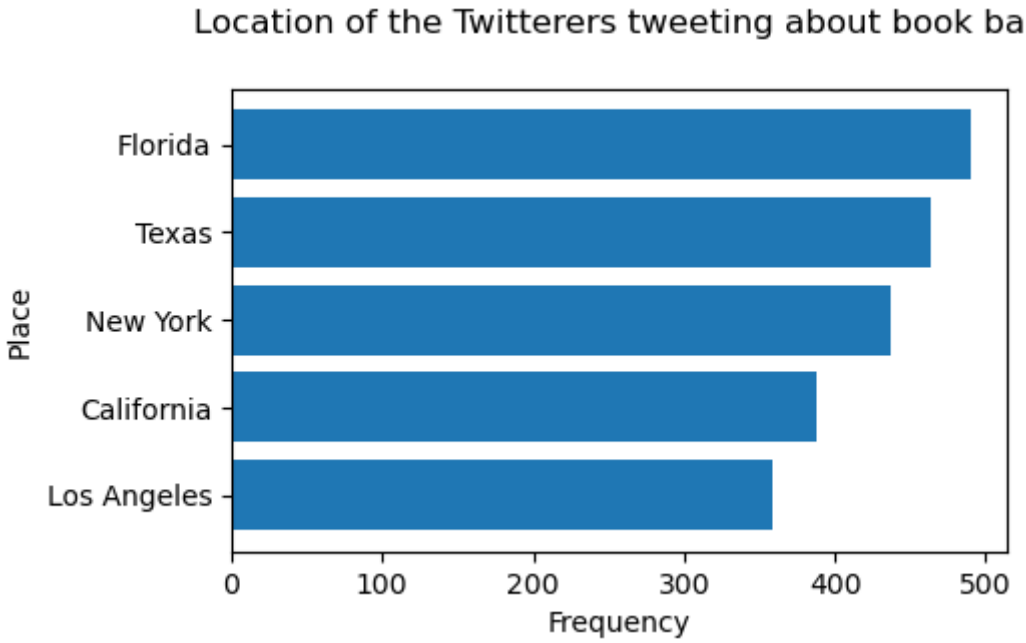
In [197…

```
loc2
```

Out[197…

	user_place	count
0	Florida	490
1	Texas	464
2	New York	437
3	California	388
4	Los Angeles	358

In [202…
```
loc2= loc2.sort_values('count')

plt.figure(figsize=(5,3))
# bar plot with matplotlib
plt.barh('user_place', 'count',data=loc2)
plt.xlabel("Frequency")
plt.ylabel("Place")
plt.title("Location of the Twitterers tweeting about book ban", size=12)
plt.show()
```



In [ ]:

In [199…
```
loc3=spark.sql("select user_place, count(*) as count \
from tweets_table2 \
where user_place is not null and user_place <> 'United States' and user_place
group by user_place \
order by count desc limit 6").toPandas()
```

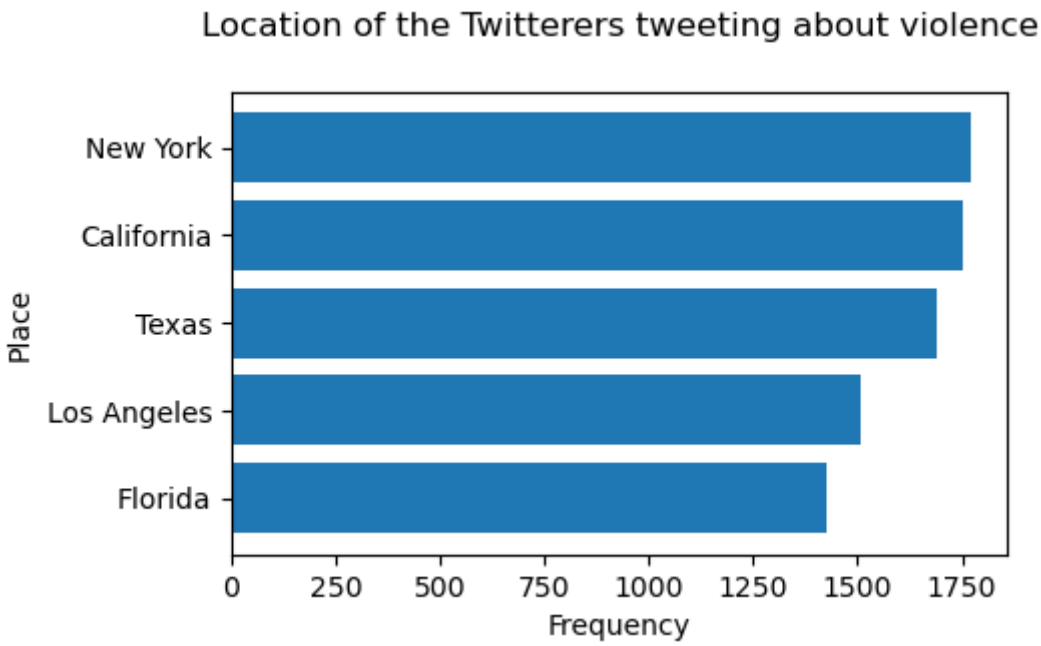
In [205…
```
loc3
```

Out[205…

	user_place	count
1	New York	1771
2	California	1750
3	Texas	1688
4	Los Angeles	1507
5	Florida	1424

In [206…
```
loc3= loc3.sort_values('count')

plt.figure(figsize=(5,3))
# bar plot with matplotlib
plt.barh('user_place', 'count',data=loc3)
plt.xlabel("Frequency")
plt.ylabel("Place")
```

```
plt.title("Location of the Twitterers tweeting about violence", size=12)
plt.show()
```



Question 3

What are the timelines of these tweets? Do you see significant peaks and valleys?

- Do you see any data collection gaps?
- Plot the timelines of these tweets

In [211…
```
import datetime

timestamp = 459286.17
dt_object = datetime.datetime.fromtimestamp(timestamp)
dt_object
```

Out[211… `datetime.datetime(1970, 1, 6, 7, 34, 46, 170000)`

In [22]:
```
from pyspark.sql import types as T
from pyspark.sql import functions as F

x=tweets_key.withColumn('as_date', F.from_unixtime((F.col('timestamp_ms')/100
```

In [18]:
```
x.select('as_date','timestamp_ms').limit(2)
```

Out[18]:

as_date	timestamp_ms
2022-05-24 22:09:56	1653430196731
2022-05-24 22:09:56	1653430196475

In [28]:
```
x=x.withColumn("as_date2", F.to_utc_timestamp(F.from_unixtime(F.col("timestam
```

In [26]:
```
x.select('as_date','as_date2','timestamp_ms').limit(2)
```

Out[26]:

as_date	as_date2	timestamp_ms

```
2022-05-29 13:46:47   2022-05-29 05:00:00   1653832007908

2022-05-29 13:46:47   2022-05-29 05:00:00   1653832007922
```

In [ ]:
```
x.write.mode("overwrite").saveAsTable("tweets_table")
```

```
22/12/09 04:44:34 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670550927691_0003_01_000060 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-09 04:44:34.572]Container killed on request. Exit co
de is 143
[2022-12-09 04:44:34.590]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.591]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670550927691_0003_01_000058 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-09 04:44:34.572]Container killed on request. Exit co
de is 143
[2022-12-09 04:44:34.573]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.590]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.deploy.yarn.YarnAllocator: Cannot find
executorId for container: container_1670550927691_0003_01_000077
22/12/09 04:44:34 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 57 on hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-stud
ents.internal: Container from a bad node: container_1670550927691_0003_01_0000
58 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-student
s.internal. Exit status: 143. Diagnostics: [2022-12-09 04:44:34.572]Container
killed on request. Exit code is 143
[2022-12-09 04:44:34.573]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.590]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 65
3.0 in stage 22.0 (TID 44077) (hub-msca-bdp-dphub-students-snigdag0402-sw-wl1
m.c.msca-bdp-students.internal executor 57): ExecutorLostFailure (executor 57
exited caused by one of the running tasks) Reason: Container from a bad node:
container_1670550927691_0003_01_000058 on host: hub-msca-bdp-dphub-students-sn
igdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 04:44:34.572]Container killed on request. Exit code is 143
[2022-12-09 04:44:34.573]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.590]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 65
9.0 in stage 22.0 (TID 44083) (hub-msca-bdp-dphub-students-snigdag0402-sw-wl1
m.c.msca-bdp-students.internal executor 57): ExecutorLostFailure (executor 57
exited caused by one of the running tasks) Reason: Container from a bad node:
container_1670550927691_0003_01_000058 on host: hub-msca-bdp-dphub-students-sn
igdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 04:44:34.572]Container killed on request. Exit code is 143
[2022-12-09 04:44:34.573]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.590]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 59 for reason Con
tainer from a bad node: container_1670550927691_0003_01_000060 on host: hub-ms
ca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit s
tatus: 143. Diagnostics: [2022-12-09 04:44:34.572]Container killed on request.
Exit code is 143
[2022-12-09 04:44:34.590]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.591]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 57 for reason Con
tainer from a bad node: container_1670550927691_0003_01_000058 on host: hub-ms
ca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit s
tatus: 143. Diagnostics: [2022-12-09 04:44:34.572]Container killed on request.
Exit code is 143
```

```
[2022-12-09 04:44:34.573]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.590]Killed by external signal
.
22/12/09 04:44:34 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 59 on hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-stud
ents.internal: Container from a bad node: container_1670550927691_0003_01_0000
60 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-wl1m.c.msca-bdp-student
s.internal. Exit status: 143. Diagnostics: [2022-12-09 04:44:34.572]Container
killed on request. Exit code is 143
[2022-12-09 04:44:34.590]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.591]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 65
0.0 in stage 22.0 (TID 44074) (hub-msca-bdp-dphub-students-snigdag0402-sw-wl1
m.c.msca-bdp-students.internal executor 59): ExecutorLostFailure (executor 59
exited caused by one of the running tasks) Reason: Container from a bad node:
container_1670550927691_0003_01_000060 on host: hub-msca-bdp-dphub-students-sn
igdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 04:44:34.572]Container killed on request. Exit code is 143
[2022-12-09 04:44:34.590]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.591]Killed by external signal
.
22/12/09 04:44:34 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 64
3.0 in stage 22.0 (TID 44067) (hub-msca-bdp-dphub-students-snigdag0402-sw-wl1
m.c.msca-bdp-students.internal executor 59): ExecutorLostFailure (executor 59
exited caused by one of the running tasks) Reason: Container from a bad node:
container_1670550927691_0003_01_000060 on host: hub-msca-bdp-dphub-students-sn
igdag0402-sw-wl1m.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 04:44:34.572]Container killed on request. Exit code is 143
[2022-12-09 04:44:34.590]Container exited with a non-zero exit code 143.
[2022-12-09 04:44:34.591]Killed by external signal
.
22/12/09 04:46:37 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 75 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000076 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-jm37.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 04:46:37 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 74 for reason Con
tainer marked as failed: container_1670550927691_0003_01_000075 on host: hub-m
sca-bdp-dphub-students-snigdag0402-sw-jm37.c.msca-bdp-students.internal. Exit
status: -100. Diagnostics: Container released on a *lost* node.
22/12/09 04:46:37 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 75 on hub-msca-bdp-dphub-students-snigdag0402-sw-jm37.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
076 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-jm37.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
22/12/09 04:46:37 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 74 on hub-msca-bdp-dphub-students-snigdag0402-sw-jm37.c.msca-bdp-stud
ents.internal: Container marked as failed: container_1670550927691_0003_01_000
075 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-jm37.c.msca-bdp-studen
ts.internal. Exit status: -100. Diagnostics: Container released on a *lost* no
de.
```

In [35]:

```
timestamp=spark.sql(" select SUBSTRING(as_date, 1, 10) as date, count(*) as c
from tweets_table \
group by SUBSTRING(as_date, 1, 10)").toPandas()
timestamp.head()
```

Out[35]:

	date	count
0	2022-10-05	403641

	date	count
1	2022-10-07	376801
2	2022-05-17	358804
3	2022-09-03	385071
4	2022-07-08	289245

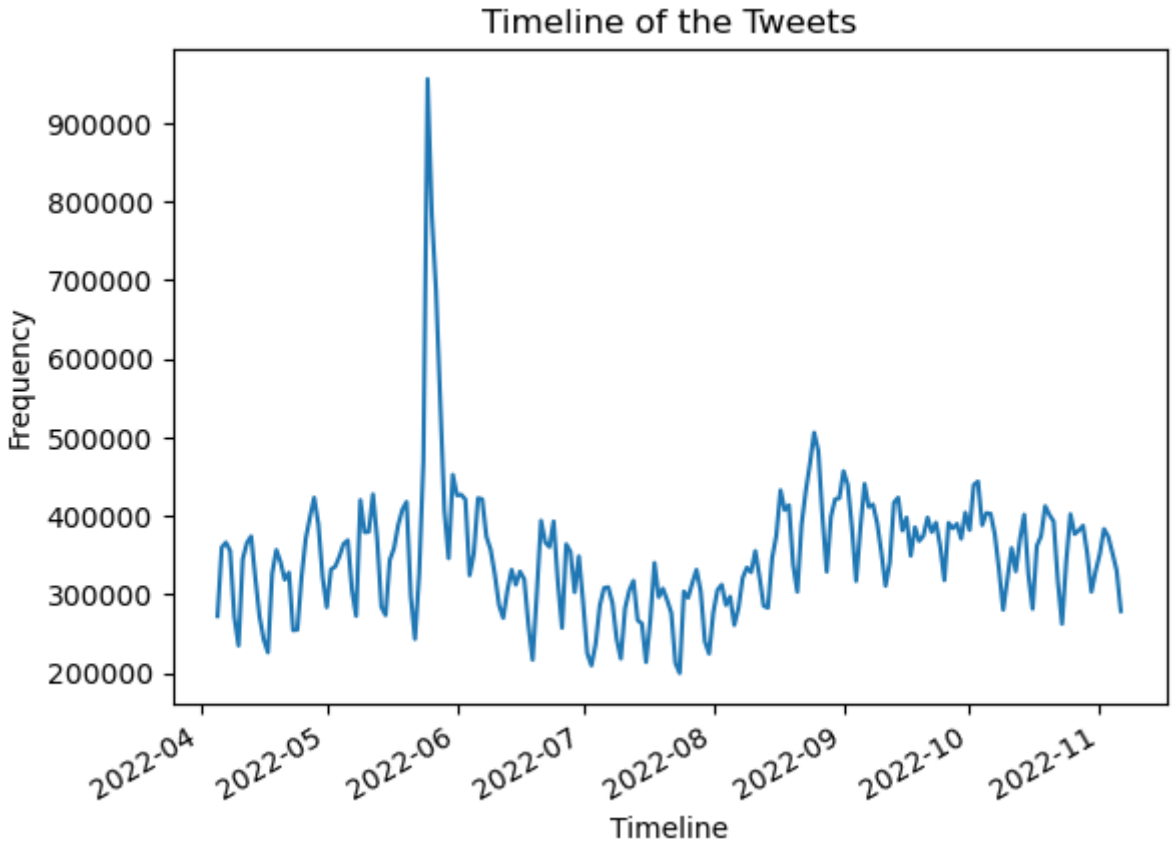
In [39]:
```
spark.sql(" select distinct SUBSTRING(as_date, 1, 4) as date \
from tweets_table ").toPandas()
```

Out[39]:

	date
0	2022

In [40]:
```
timestamp=timestamp.sort_values(by=['date'])
```

In [51]:
```
import pandas as pd
import matplotlib.pyplot as plt

#date_time = ["2021-01-01", "2021-01-02", "2021-01-03"]
date_time = pd.to_datetime(list(timestamp["date"]))
data = list(timestamp["count"])

DF = pd.DataFrame()
DF['value'] = data
DF = DF.set_index(date_time)
plt.plot(DF)
plt.gcf().autofmt_xdate()

plt.xlabel("Timeline")
plt.ylabel("Frequency")
plt.title("Timeline of the Tweets", size=12)

plt.show()
```



In [ ]:

Question 4

In [3]:
```
import pandas as pd
import numpy as np
pd.set_option('display.max_colwidth', None)
pd.reset_option('display.max_rows')
from itertools import compress
from pyspark.sql.functions import *
from pyspark.sql.types import *
import seaborn as sns
import matplotlib.pyplot as plt
warnings.filterwarnings(action='ignore')
```

In [5]:
```
import re
from pyspark.ml.feature import MinHashLSH
from pyspark.ml.feature import CountVectorizer,  IDF, CountVectorizerModel, T
from pyspark import SparkContext
from pyspark.sql import SQLContext
from pyspark.sql import Row
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
```

In [21]:
```
df_text_raw = tweets_key.select('text').limit(30000)
```

In [22]:
```
df_text_raw.limit(5)
```

Out[22]:

text
RT @ABC: “Why are...

Indiana High Scho...

RT @jaketapper: F...

#Uvalde is just a...

RT @Josh_Moon: 14...

Step 1. Clean the data, remove stopwords and create index

In [27]:
```
text = df_text_raw.rdd.map(lambda x : x['text']).filter(lambda x: x is not No

#StopWords = stopwords.words("english")

tokens = text\
    .map( lambda document: document.strip().lower())\
    .map( lambda document: re.split(" ", document))\
    .map( lambda word: [x for x in word if x.isalnum()])\
    .map( lambda word: [x for x in word if len(x) > 3] )
    #.map( lambda word: [x for x in word if x not in StopWords])\
    #.zipWithIndex()
```

In [28]:
```
row = Row('text')
df_text = text.map(row).zipWithIndex().toDF(['text','id'])
df_text.limit(5)
```

```
22/12/09 02:16:50 WARN org.apache.spark.deploy.yarn.YarnAllocator: Container f
rom a bad node: container_1670550927691_0001_01_000001 on host: hub-msca-bdp-d
phub-students-snigdag0402-sw-5270.c.msca-bdp-students.internal. Exit status: 1
43. Diagnostics: [2022-12-09 02:16:50.610]Container killed on request. Exit co
de is 143
[2022-12-09 02:16:50.611]Container exited with a non-zero exit code 143.
[2022-12-09 02:16:50.611]Killed by external signal
.
22/12/09 02:16:50 ERROR org.apache.spark.scheduler.cluster.YarnScheduler: Lost
executor 1 on hub-msca-bdp-dphub-students-snigdag0402-sw-5270.c.msca-bdp-stude
nts.internal: Container from a bad node: container_1670550927691_0001_01_00000
1 on host: hub-msca-bdp-dphub-students-snigdag0402-sw-5270.c.msca-bdp-student
s.internal. Exit status: 143. Diagnostics: [2022-12-09 02:16:50.610]Container
killed on request. Exit code is 143
[2022-12-09 02:16:50.611]Container exited with a non-zero exit code 143.
[2022-12-09 02:16:50.611]Killed by external signal
.
22/12/09 02:16:50 WARN org.apache.spark.scheduler.cluster.YarnSchedulerBackend
$YarnSchedulerEndpoint: Requesting driver to remove executor 1 for reason Cont
ainer from a bad node: container_1670550927691_0001_01_000001 on host: hub-msc
a-bdp-dphub-students-snigdag0402-sw-5270.c.msca-bdp-students.internal. Exit st
atus: 143. Diagnostics: [2022-12-09 02:16:50.610]Container killed on request.
Exit code is 143
[2022-12-09 02:16:50.611]Container exited with a non-zero exit code 143.
[2022-12-09 02:16:50.611]Killed by external signal
.
22/12/09 02:16:50 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 22
94.0 in stage 4.0 (TID 17697) (hub-msca-bdp-dphub-students-snigdag0402-sw-527
0.c.msca-bdp-students.internal executor 1): ExecutorLostFailure (executor 1 ex
ited caused by one of the running tasks) Reason: Container from a bad node: co
ntainer_1670550927691_0001_01_000001 on host: hub-msca-bdp-dphub-students-snig
dag0402-sw-5270.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 02:16:50.610]Container killed on request. Exit code is 143
[2022-12-09 02:16:50.611]Container exited with a non-zero exit code 143.
[2022-12-09 02:16:50.611]Killed by external signal
.
22/12/09 02:16:50 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 22
90.0 in stage 4.0 (TID 17693) (hub-msca-bdp-dphub-students-snigdag0402-sw-527
0.c.msca-bdp-students.internal executor 1): ExecutorLostFailure (executor 1 ex
ited caused by one of the running tasks) Reason: Container from a bad node: co
ntainer_1670550927691_0001_01_000001 on host: hub-msca-bdp-dphub-students-snig
```

```
dag0402-sw-5270.c.msca-bdp-students.internal. Exit status: 143. Diagnostics:
[2022-12-09 02:16:50.610]Container killed on request. Exit code is 143
[2022-12-09 02:16:50.611]Container exited with a non-zero exit code 143.
[2022-12-09 02:16:50.611]Killed by external signal
.
```

Out[28]:

	text	id
	{RT @21savage: I ...	0
	{RT @MoneyMiaaaa:...	1
	{RT @ABC7: #BREAK...	2
	{RT @Rajoo_Bhau: ...	3
	{RT @loeytaro: he...	4

In [42]:
```
df_tokens = spark.createDataFrame(tokens, ["list_of_words",'id'])

#Drop records with no tokens
df_tokens = df_tokens.where(col('list_of_words').getItem(0).isNotNull())
```

Step 2. Fit countvectorizer to create word features

In [20]:
```
vectorize = CountVectorizer(inputCol="list_of_words", outputCol="features", m
df_vectorize = vectorize.fit(df_tokens).transform(df_tokens)
```

```
22/12/09 00:20:11 WARN org.apache.spark.scheduler.TaskSetManager: Lost task 0.
9 in stage 29.0 (TID 20833) (hub-msca-bdp-dphub-students-snigdag0402-w-1.c.msc
a-bdp-students.internal executor 13): org.apache.spark.SparkException: Failed
to execute user defined function(LSHModel$Lambda$3726/1460921732: (struct<typ
e:tinyint,size:int,indices:array<int>,values:array<double>>) => array<struct<t
ype:tinyint,size:int,indices:array<int>,values:array<double>>>)
        at org.apache.spark.sql.catalyst.expressions.GeneratedClass$GeneratedI
teratorForCodegenStage1.processNext(Unknown Source)
        at org.apache.spark.sql.execution.BufferedRowIterator.hasNext(Buffered
RowIterator.java:43)
        at org.apache.spark.sql.execution.WholeStageCodegenExec$$anon$1.hasNex
t(WholeStageCodegenExec.scala:755)
        at org.apache.spark.sql.execution.columnar.DefaultCachedBatchSerialize
r$$anon$1.next(InMemoryRelation.scala:87)
        at org.apache.spark.sql.execution.columnar.DefaultCachedBatchSerialize
r$$anon$1.next(InMemoryRelation.scala:79)
        at scala.collection.Iterator$$anon$10.next(Iterator.scala:461)
        at org.apache.spark.storage.memory.MemoryStore.putIterator(MemoryStor
e.scala:222)
        at org.apache.spark.storage.memory.MemoryStore.putIteratorAsValues(Mem
oryStore.scala:299)
        at org.apache.spark.storage.BlockManager.$anonfun$doPutIterator$1(Bloc
kManager.scala:1425)
        at org.apache.spark.storage.BlockManager.org$apache$spark$storage$Bloc
kManager$$doPut(BlockManager.scala:1352)
        at org.apache.spark.storage.BlockManager.doPutIterator(BlockManager.sc
ala:1416)
        at org.apache.spark.storage.BlockManager.getOrElseUpdate(BlockManager.
scala:1239)
        at org.apache.spark.rdd.RDD.getOrCompute(RDD.scala:384)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:335)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
```

```
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.rdd.MapPartitionsRDD.compute(MapPartitionsRDD.scal
a:52)
        at org.apache.spark.rdd.RDD.computeOrReadCheckpoint(RDD.scala:373)
        at org.apache.spark.rdd.RDD.iterator(RDD.scala:337)
        at org.apache.spark.shuffle.ShuffleWriteProcessor.write(ShuffleWritePr
ocessor.scala:59)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:99)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:52)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Execut
or.scala:505)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:50
8)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:750)
Caused by: java.lang.IllegalArgumentException: requirement failed: Must have a
t least 1 non zero entry.
        at scala.Predef$.require(Predef.scala:281)
        at org.apache.spark.ml.feature.MinHashLSHModel.hashFunction(MinHashLS
H.scala:61)
        at org.apache.spark.ml.feature.LSHModel.$anonfun$transform$1(LSH.scal
a:101)
        ... 39 more

22/12/09 00:20:11 ERROR org.apache.spark.scheduler.TaskSetManager: Task 0 in s
tage 29.0 failed 10 times; aborting job
```

In [21]:

```
df_vectorize.limit(5).toPandas()
```

```
22/12/09 00:20:30 WARN org.apache.spark.scheduler.DAGScheduler: Broadcasting l
arge task binary with size 1682.4 KiB
```

Out[21]:

	list_of_words	id	features
0	[school, thot, getting, pounded]	0	(1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...)
1	[students, learn, practice, critical, thinking, color]	1	(0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...)

	list_of_words	id	features
2	[schools, philippines, focus, bullying, concerns, take, rather, focusing]	2	(0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...)
3	[idle, singing, seduced, lesbian, demon, cute, school, outfits]	3	(1.0, 0.0, ...)
4	[small, account, savings, school, tuition]	4	(1.0, 0.0, ...)

Step 3. Fit MinHashLSH to create hash table

Note: Adding more hash tables will increase the accuracy at the expense of training time

In [23]:
```
mh = MinHashLSH(inputCol="features", outputCol="hashes", numHashTables=5)
model = mh.fit(df_vectorize)
df_hashed = mh.fit(df_vectorize).transform(df_vectorize).cache()
```

In [24]:
```
df_hashed_text = df_text.join(df_hashed, "id", how = 'left').cache()

df_hashed_text.limit(5)
```

```
22/12/09 00:26:00 WARN org.apache.spark.scheduler.DAGScheduler: Broadcasting l
arge task binary with size 1708.8 KiB
22/12/09 00:27:34 WARN org.apache.spark.scheduler.DAGScheduler: Broadcasting l
arge task binary with size 1749.1 KiB
22/12/09 00:27:54 WARN org.apache.spark.scheduler.DAGScheduler: Broadcasting l
arge task binary with size 1749.1 KiB
```

Out[24]:

id	text	list_of_words	features	hashes
26	{RT @SuperYakiSho...	[gloating, encour...	(158763,[1880,223...	[[1.065303651E9],...
29	{RT @serineblack:...	[florida, high, s...	(158763,[0,2,21,4...	[[5.6714553E7], [...
474	{Your weekly page...	[karnataka, paint...	(158763,[0,76,204...	[[9.4237756E7], [...
964	{why does mark lo...	[anyone, else, pa...	(158763,[0,36,38,...	[[4510477.0], [3....
1677	{Velma-Alma vs Ce...	[pretty, sure, sc...	(158763,[4,182,19...	[[2.8431946E7], [...

Step 4. Establish similarity threshold and return near-duplicate records

Note: we are joining dataframe to itself to get near-duplicate pairs

In [27]:
```
df_hashed_text.write.mode("overwrite").saveAsTable("df_hashed_text_table")
```

```
ivysettings.xml file not found in HIVE_HOME or HIVE_CONF_DIR,/etc/hive/conf.di
```

```
st/ivysettings.xml will be used
22/12/09 00:37:09 WARN org.apache.spark.scheduler.DAGScheduler: Broadcasting l
arge task binary with size 1941.2 KiB
22/12/09 00:37:35 WARN org.apache.hadoop.hive.ql.session.SessionState: METASTO
RE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is s
et to instance of HiveAuthorizerFactory.
```

In [25]:
```
jaccard_distance = 0.5

df_dups_text = model.approxSimilarityJoin(df_hashed_text, df_hashed_text, jac
            col("distCol"),
            col("datasetA.id").alias("id_A"),
            col("datasetB.id").alias("id_B"),
            col('datasetA.text').alias('text_A'),
            col('datasetB.text').alias('text_B'),
#             col('datasetA.list_of_words').alias('words_A'),
#             col('datasetB.list_of_words').alias('words_B')
            )
```

In [26]:
```
df_dups_50 = df_dups_text
```

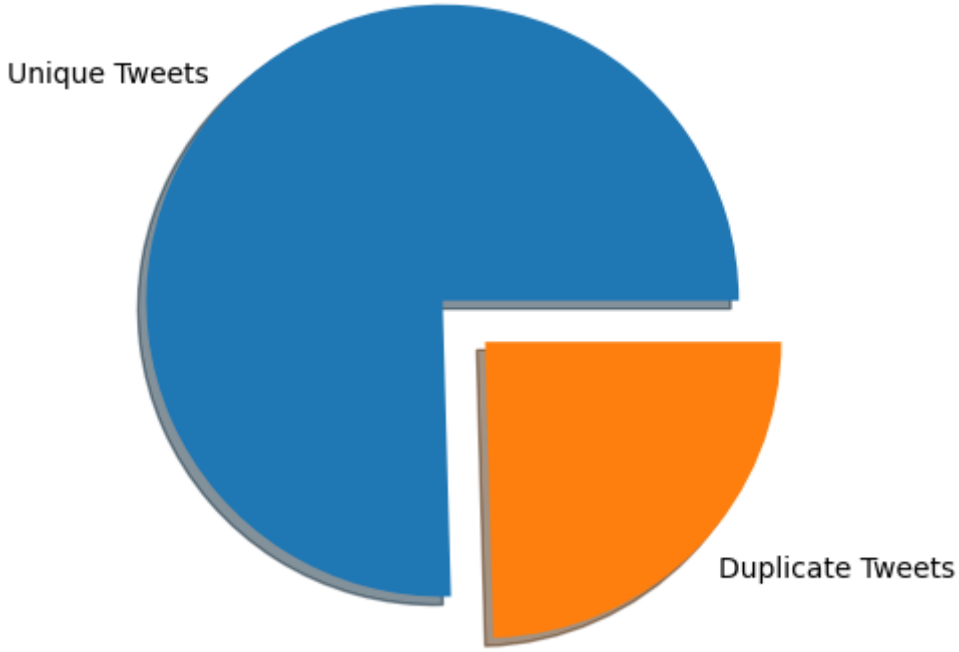
In [28]:
```
records=spark.sql('select count(*) from df_hashed_text_table;').toPandas()
```

In [30]:
```
r=records.iloc[0,0]
```

In [47]:
```
records = r
d=df_dups_50.select('id_A').distinct().count()
uniques = r-d

print ('Percentage of Duplicate tweets based on {', jaccard_distance, '} jacc
print ('Percentage of Unique tweets based on {', jaccard_distance, '} jaccard
```

```
Percentage of Duplicate tweets based on { 0.5 } jaccard distance:  0.5 : 24.5
4%
Percentage of Unique tweets based on { 0.5 } jaccard distance:  0.5 : 75.45%
```

In [50]:
```
import matplotlib.pyplot as plt
import numpy as np

y = np.array([75.45, 24.55])
mylabels = ["Unique Tweets", "Duplicate Tweets"]
myexplode = [0.2, 0]

plt.pie(y, labels = mylabels, explode = myexplode, shadow = True)
plt.show()
```



In [ ]:

In [ ]: