

PES UNIVERSITY

100 feet Ring Road, BSK 3rd Stage
Bengaluru 560085



Department of Computer Science and Engineering
B. Tech. CSE - 6th Semester
Jan – May 2022

UE19CS344
DATABASE TECHNOLOGIES (DBT)

Project Report

Crime data analysis

PES2UG19CS013: Achyut Jagini
PES2UG19CS230: Melavoy Nithin Reddy
PES2UG19CS189: Koduru Bharath Subba Reddy

Table of Contents

1. Introduction
2. Installation of Software
3. Input Data
 1. Source
 2. Description
4. Streaming Mode Experiment
 1. Description
 2. Windows
 3. Results
5. Batch mode Experiment
 1. Description
 2. Data Size
 3. Results
6. Comparison of Streaming vs Batch mode Results
7. Conclusion
8. References

Introduction

This project aims to perform analysis of crime data using spark and kafka.

Installation of software

Software installed and used are

Spark

Kafka

Zookeeper

Kafka Configuration Manager(CMAK)

Mongodb database

Input data

Source

For spark streaming ,data input from crime-train.csv and test.csv datasets.

For kafka ,data produced and consumed
police-department-calls-for-service.json

Description

Train.csv has columns Dates ,Category, Descript, DayOfWeek,PdDistrict,Resolution,Address,X,Y.

police-department-calls-for-service.json

Is json data with format

"crime_id": "183653763",

"original_crime_type_name": "Traffic Stop",

"report_date": "2018-12-31T00:00:00.000",

"call_date": "2018-12-31T00:00:00.000",

"offense_date": "2018-12-31T00:00:00.000",

"call_time": "23:57",
"call_date_time": "2018-12-31T23:57:00.000",
"disposition": "ADM",
"address": "Geary Bl/divisadero St",
"city": "San Francisco",
"state": "CA",
"agency_id": "1",
"address_type": "Intersection","common_location": ""

Real time processing

Total time-1220:33:35 5:00s/it

```
pes2ug19cs013@pes2ug19cs013-VirtualBox:~/Desktop/proj$ python3 stream.py -f crime -b 1
Namespace(batch_size=1, endless=False, file='crime')
Waiting for connection on port 6100...
Connected to ('127.0.0.1', 57714)
Starting to stream crime dataset
0%| | 55/878050 [04:35<1220:39:25, 5.01s/it]
```

```
pes2ug19cs013@pes2ug19cs013-VirtualBox:~/Desktop/proj$ /opt/spark/bin/spark-submit client.py 2>log.txt
+-----+
|      Dates|Category|      Descript|DayOfWeek|PdDistrict|      Resolution|      Address|      X|      Y|
+-----+
|2015-05-13 23:53:00|WARRANTS|WARRANT ARREST|Wednesday|  NORTHERN|ARREST, BOOKED|OAK ST / LAGUNA ST|-122.425891675136|37.7745985956747|
+-----+
+-----+
|Category|count|
+-----+
|WARRANTS|  1|
+-----+
+-----+
|      Descript|count|
+-----+
|WARRANT ARREST|  1|
+-----+
+-----+
|first_record_date|latest_record_date|
+-----+
|2015-05-13 23:53:00|2015-05-13 23:53:00|
+-----+
+-----+
|      Dates|      date_time|      month|
+-----+
|2015-05-13 23:53:00|2015-05-13 23:53:00|2015-01-01|
+-----+
+-----+
|      Resolution|      month|count|
+-----+
|ARREST, BOOKED|2015-01-01|  1|
+-----+
+-----+
|Resolution|      month|count|yearpd|arrest|
+-----+
|0 ARREST, BOOKED|2015-01-01|  1|2015|True|
+-----+
|hour|count|
+-----+
|0|23|1|
+-----+
+-----+
|PdDistrict|count|
+-----+
|NORTHERN|  1|
+-----+
```

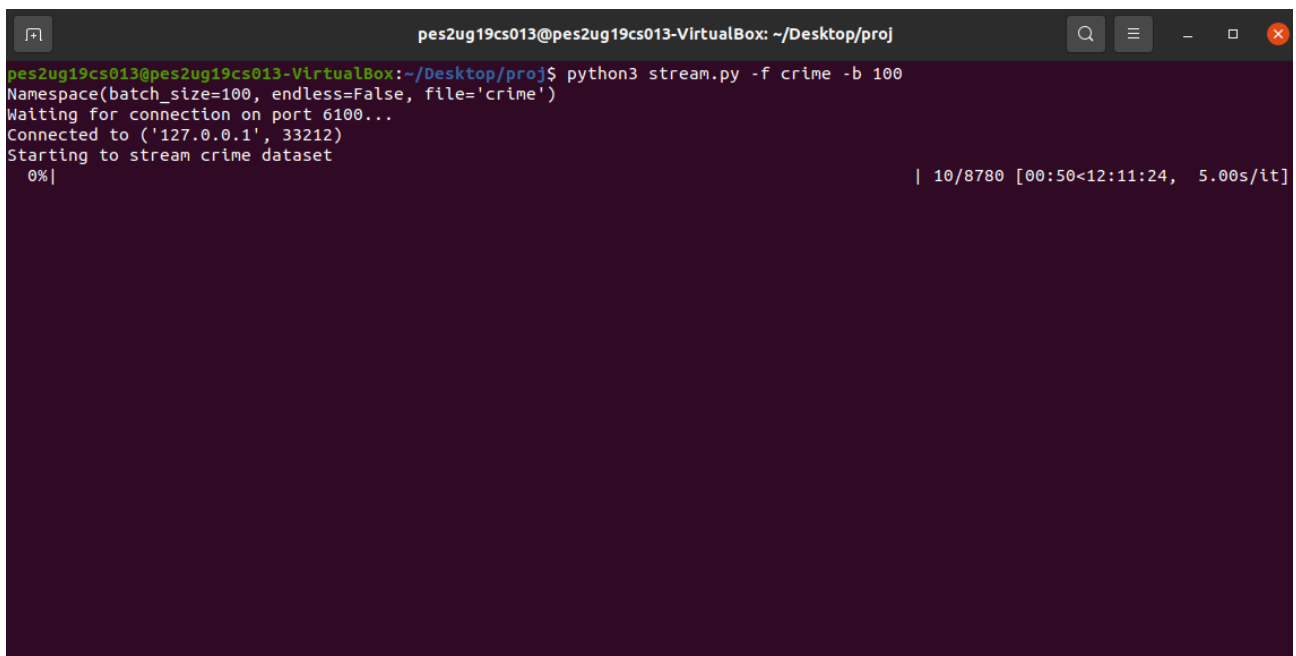
Batch Streaming experiment

Data streamed using stream.py file

python3 stream.py -f crime -b 100(can vary 100 window size)

Total time-12:11:24 5.00 s/iter

window size 100



```
pes2ug19cs013@pes2ug19cs013-VirtualBox: ~/Desktop/proj
pes2ug19cs013@pes2ug19cs013-VirtualBox:~/Desktop/proj$ python3 stream.py -f crime -b 100
Namespace(batch_size=100, endless=False, file='crime')
Waiting for connection on port 6100...
Connected to ('127.0.0.1', 33212)
Starting to stream crime dataset
0%|                                     | 10/8780 [00:50<12:11:24, 5.00s/it]
```

client.py

/opt/spark/bin/spark-submit client.py 2>log.txt

Code

```
1 # importing required libraries
2 #importing all directories needed
3 import os
4 import sys, pyspark, json
5 from pyspark import SparkContext
6 import numpy as np
7 import csv
8 import matplotlib.pyplot as plt
9
10 import pyspark.sql.types as tp
11
12 from pyspark.sql import SQLContext, SparkSession
13 from pyspark.sql import SparkSession, Row, Column
14 from pyspark.streaming import StreamingContext
15
16 from pyspark.sql.functions import col
17
18 from pyspark.sql.functions import *
19
20 from pyspark.sql.types import FloatType
21
22 import pyspark.sql.functions as F
23 from pyspark.sql.types import IntegerType
24 from pyspark.sql.functions import lit, monotonically_increasing_id, row_number
25 from pyspark.sql import Window
26 import datetime
27 import pandas as pd
28
29
30
31 sc = SparkContext("local[2]", "Crime")
32 #spark = SparkSession(sc)
33 ssc = StreamingContext(sc, 1)
34 sql_context=SQLContext(sc)
35
36
37 def dfto(data):
38     if data.isEmpty():
39         return
40     s = SparkSession(data.context)
41     data = data.collect()[0]
42     #converting data into dataframe for easy usage
43     cols = [f'Feature{j}' for j in range(len(data[0]))]
44     colm = ['Dates', 'Category', 'Descript', 'DayOfWeek', 'PdDistrict', 'Resolution', 'Address', 'X', 'Y']
45     df = s.createDataFrame(data, colm)
```



```

46 df.show()
47 df = df.withColumn('X', df['X'].cast(FloatType()))
48 df = df.withColumn('Y', df['Y'].cast(FloatType()))
49
50
51 df.groupBy("Category").count().orderBy(col("count").desc()).show()
52
53 df.groupBy("Descript").count().orderBy(col("count").desc()).show()
54
55 df.select(min('Dates').alias('first_record_date'),max('Dates').alias('latest_record_date')).show(truncate=False)
56
57
58 df = df.withColumn('date_time', to_timestamp('Dates', 'yyyy-MM-dd HH:mm:ss'))\
59 .withColumn('month', trunc('date_time', 'YYYY')) #adding a month column to be able to view stats on a monthly basis
60
61 df.select(['Dates', 'date_time', 'month'])\
62 .show()
63
64
65 type_arrest_date = df.groupBy(['Resolution', 'month'])\
66 .count()\
67 .orderBy(['month', 'count'])
68 print()
69 type_arrest_date.show()
70
71
72 type_arrest_pddf = pd.DataFrame(type_arrest_date.rdd.map(lambda l: l.asDict()).collect())
73
74
75 type_arrest_pddf['yearpd'] = type_arrest_pddf['month'].apply(lambda dt: datetime.datetime.strptime(pd.Timestamp(dt), '%Y'))
76 #type_arrest_pddf['arrest'] = type_arrest_pddf['arrest'].apply(lambda l: l=='True')
77 type_arrest_pddf['arrest'] = type_arrest_pddf['Resolution'].apply(lambda l: "ARREST" in l)
78 print(type_arrest_pddf)
79
80
81 #What time of the day are cimnal the busiest?
82
83 df_hour = df.withColumn('hour', hour(df['date_time']))
84 hourly_count = df_hour.groupBy('hour').count().cache()
85 hourly_total_count = hourly_count.groupBy('hour').sum('count')
86

```

```

86
87 hourly_count_pddf = pd.DataFrame(hourly_total_count.select(hourly_total_count['hour'], hourly_total_count['sum(count)'].alias('count'))\
88 .rdd.map(lambda l: l.asDict()).\
89 .collect())
90
91 print(hourly_count_pddf)
92
93
94 #df.select('PdDistrict').distinct().count()
95
96 #What are the top disctricts where crime occurred?
97
98 df.groupBy(['PdDistrict']).count().orderBy('count', ascending=False).show()
99
100 #street_home_hour = location_hour.where((location_hour['PdDistrict'] == 'CENTRAL') | (location_hour['PdDistrict'] == 'NORTHERN'))
101 # data frame with location descriptions and counts of recorded crimes, and hours...
102 #street_home_hour_pddf = pd.DataFrame(street_home_hour.rdd.map(lambda row: row.asDict()).collect())
103 #street_home_hour_pddf = street_home_hour_pddf.sort_values(by='hour')
104
105 #print(street_home_hour_pddf)
106
107 categorydf = pd.DataFrame(df_hour.groupBy(['Category', 'hour']).count().orderBy('hour').rdd.map(lambda row: row.asDict()).collect())
108 print(categorydf)
109
110
111 #dom = domestic_hour[domestic_hour['Category==NON-CRIMINAL'] == 'True']['count']
112 #non_dom = domestic_hour[domestic_hour['Category==NON-CRIMINAL'] == 'False']['count']
113
114 #either_dom = domestic_hour.groupby(by=['hour']).sum()['count']
115
116 #dom_keys = domestic_hour[domestic_hour['domestic'] == 'False']['hour']
117
118 df_dates = df_hour.withColumn('week_day', dayofweek(df_hour['date_time']))\
119 .withColumn('year_month', month(df_hour['date_time']))\
120 .withColumn('month_day', dayofmonth(df_hour['date_time']))\
121 .withColumn('date_number', datediff(df['date_time'], to_date(lit('2001-01-01'), format='yyyy-MM-dd')))\
122 .cache()
123
124 df_dates.select(['Dates', 'month', 'hour', 'week_day', 'year_month', 'month_day', 'date_number']).show(20, truncate=False)
125
126 week_day_crime_counts = df_dates.groupBy('week_day').count()
127 week_day_crime_counts_pddf = pd.DataFrame(week_day_crime_counts.orderBy('week_day').rdd.map(lambda e: e.asDict()).collect())
128 print(week_day_crime_counts_pddf)
129
130

```

```

131
132 month_day_crime_counts = df_dates.groupBy('month_day').count()
133 month_day_crime_counts_pddf = pd.DataFrame(month_day_crime_counts.orderBy('month_day')).rdd.map(lambda e: e.asDict()).collect()
134 month_day_crime_counts_pddf.sort_values(by='count', ascending=False).head(10)
135
136
137 month_day_crime_counts_pddf = month_day_crime_counts_pddf.sort_values(by='month_day', ascending=True)
138 print(month_day_crime_counts_pddf)
139
140
141 df_dates_community_areas = df_dates.na.drop(subset=['address']).groupBy('address').count()
142 df_dates_community_areas.orderBy('count', ascending=False).show(10)
143
144 top_crime_types = df_dates.select('Category').groupBy('Category').count().rdd.map(lambda row: row.asDict()).takeOrdered(10, key=lambda l: 1/l['count'])
145
146 top_busy_areas = df_dates_community_areas.rdd.map(lambda row: row.asDict()).takeOrdered(10, key=lambda l: 1/l['count'])
147
148 top_crime_types_lst = [dc['Category'] for dc in top_crime_types]
149 top_busy_areas_lst = [dc['address'] for dc in top_busy_areas]
150
151
152
153 print(top_crime_types_lst)
154 print(top_busy_areas_lst)
155
156
157
158 def mapdat(data):
159     lstrecord = list()
160     js_dat=json.loads(data)
161     for record in js_dat:
162         tup = tuple(js_dat[record].values())
163         lstrecord.append(tup)
164     return lstrecord
165
166 lines = ssc.socketTextStream("localhost",6100).map(mapdat).foreachRDD(dfdo)
167
168 ssc.start()
169 #ssc.awaitTermination(100)
170 #ssc.stop()
171 ssc.awaitTermination()

```

Output

```

pes2ug19cs013@pes2ug19cs013-VirtualBox: ~/Desktop/proj$ /opt/spark/bin/spark-submit client.py 2>log.txt
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Dates | Category | Description | DayOfWeek | PdDistrict | Resolution | Address | X | Y |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 2015-05-13 23:53:00 | WARRANTS | WARRANT ARREST | Wednesday | NORTHERN | ARREST, BOOKED | OAK ST / LAGUNA ST | -122.425891675136 | 37.7745985956747 |
| 2015-05-13 23:53:00 | OTHER OFFENSES | TRAFFIC VIOLATION... | Wednesday | NORTHERN | ARREST, BOOKED | OAK ST / LAGUNA ST | -122.425891675136 | 37.7745985956747 |
| 2015-05-13 23:33:00 | OTHER OFFENSES | TRAFFIC VIOLATION... | Wednesday | NORTHERN | ARREST, BOOKED | VANNES AV / GREE... | -122.42436302145 | 37.8004143219856 |
| 2015-05-13 23:30:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | NORTHERN | NONE | 1500 Block of LOW... | -122.426995326766 | 37.8008726327692 |
| 2015-05-13 23:30:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | PARK | NONE | 100 Block of BROD... | -122.438737622757 | 37.7715411720578 |
| 2015-05-13 23:30:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | INGLESIDE | NONE | 0 Block of TEDDY AV | -122.403252361212 | 37.713436704116 |
| 2015-05-13 23:30:00 | VEHICLE THEFT | STOLEN AUTOMOBILE | Wednesday | INGLESIDE | NONE | AVALON AV / PERU AV | -122.423326976668 | 37.7251380403778 |
| 2015-05-13 23:30:00 | VEHICLE THEFT | STOLEN AUTOMOBILE | Wednesday | BAYVIEW | NONE | KIRKWOOD AV / DON... | -122.371274317441 | 37.7275640719518 |
| 2015-05-13 23:00:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | RICHMOND | NONE | 600 Block of 47TH AV | -122.508194031117 | 37.7766012606812 |
| 2015-05-13 23:00:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | CENTRAL | NONE | JEFFERSON ST / LE... | -122.419087676747 | 37.8078015516515 |
| 2015-05-13 22:58:00 | LARCENY/THEFT | PETTY THEFT FROM ... | Wednesday | CENTRAL | NONE | JEFFERSON ST / LE... | -122.419087676747 | 37.8078015516515 |
| 2015-05-13 22:30:00 | OTHER OFFENSES | MISCELLANEOUS INV... | Wednesday | TARAVAL | NONE | 0 Block of ESCOLT... | -122.487983072777 | 37.7376666543327 |
| 2015-05-13 22:30:00 | VANDALISM | MALICIOUS MISCHIE... | Wednesday | TENDERLOIN | NONE | TURK ST / JONES ST | -122.412414263581 | 37.7830037964534 |
| 2015-05-13 22:06:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | NORTHERN | NONE | FILLMORE ST / GEA... | -122.432914603494 | 37.7843533426568 |
| 2015-05-13 22:00:00 | NON-CRIMINAL | FOUND PROPERTY | Wednesday | BAYVIEW | NONE | 200 Block of WILL... | -122.307744427103 | 37.7299346936044 |
| 2015-05-13 22:00:00 | NON-CRIMINAL | FOUND PROPERTY | Wednesday | BAYVIEW | NONE | 0 Block of MENDEL... | -122.383691503959 | 37.7431890419965 |
| 2015-05-13 22:00:00 | ROBBERY | ROBBERY, ARMED VI... | Wednesday | TENDERLOIN | NONE | EDDY ST / JONES ST | -122.412597377187 | 37.7839320277272 |
| 2015-05-13 21:55:00 | ASSAULT | AGGRAVATED ASSAUL... | Wednesday | INGLESIDE | NONE | CODEUS ST / MISSI... | -122.421681531572 | 37.7428222004845 |
| 2015-05-13 21:40:00 | OTHER OFFENSES | TRAFFIC VIOLATION | Wednesday | BAYVIEW | ARREST, BOOKED | MENDELL ST / HUDS... | -122.386400869953 | 37.738983491072 |
| 2015-05-13 21:30:00 | NON-CRIMINAL | FOUND PROPERTY | Wednesday | TENDERLOIN | NONE | 100 Block of JONE... | -122.412249767634 | 37.782556330202 |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

+-----+-----+
| Category | count |
+-----+-----+
| LARCENY/THEFT | 45 |
| OTHER OFFENSES | 13 |
| NON-CRIMINAL | 13 |
| VEHICLE THEFT | 6 |
| WARRANTS | 6 |
| VANDALISM | 4 |
| BURGLARY | 3 |
| ROBBERY | 3 |
| ASSAULT | 2 |
| SUSPICIOUS OCC | 2 |
| WEAPON LAWS | 2 |
| DRUNKENNESS | 1 |
+-----+-----+

```

```

[2015-05-13 23:53:00|2015-05-13 23:53:00|2015-01-01|
[2015-05-13 23:33:00|2015-05-13 23:33:00|2015-01-01|
[2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
[2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
[2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
[2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
[2015-05-13 23:00:00|2015-05-13 23:00:00|2015-01-01|
[2015-05-13 23:00:00|2015-05-13 23:00:00|2015-01-01|
[2015-05-13 22:58:00|2015-05-13 22:58:00|2015-01-01|
[2015-05-13 22:30:00|2015-05-13 22:30:00|2015-01-01|
[2015-05-13 22:30:00|2015-05-13 22:30:00|2015-01-01|
[2015-05-13 22:06:00|2015-05-13 22:06:00|2015-01-01|
[2015-05-13 22:00:00|2015-05-13 22:00:00|2015-01-01|
[2015-05-13 22:00:00|2015-05-13 22:00:00|2015-01-01|
[2015-05-13 21:55:00|2015-05-13 21:55:00|2015-01-01|
[2015-05-13 21:40:00|2015-05-13 21:40:00|2015-01-01|
[2015-05-13 21:30:00|2015-05-13 21:30:00|2015-01-01|
+-----+

```

only showing top 20 rows

```

+-----+-----+-----+
| Resolution| month|count|
+-----+-----+-----+
| PSYCHOPATHIC CASE|2015-01-01| 1|
| ARREST, CITED|2015-01-01| 2|
| ARREST, BOOKED|2015-01-01| 11|
| NONE|2015-01-01| 86|
+-----+-----+-----+

```

```

      Resolution      month      count      yearpd      arrest
0  PSYCHOPATHIC CASE  2015-01-01      1      2015      False
1  ARREST, CITED    2015-01-01      2      2015      True
2  ARREST, BOOKED   2015-01-01     11      2015      True
3      NONE         2015-01-01     86      2015      False
   hour      count
0      22          7
1      20         19
2      19         27
3      23         10
4      21         12
5      18         25
+-----+

```

```

+-----+-----+
| Descript|count|
+-----+-----+
| GRAND THEFT FROM ...| 28|
| AIDED CASE, MENTA...| 6|
| GRAND THEFT FROM ...| 5|
| PETTY THEFT FROM ...| 4|
| STOLEN AUTOMOBILE| 4|
| WARRANT ARREST| 3|
| DRIVERS LICENSE, ...| 3|
| TRAFFIC VIOLATION...| 3|
| TRAFFIC VIOLATION| 3|
| FOUND PROPERTY| 3|
| MALICIOUS MISCHIE...| 3|
| ENROUTE TO OUTSID...| 2|
| LOST PROPERTY| 2|
| GRAND THEFT FROM ...| 2|
| MISCELLANEOUS INV...| 2|
| AGGRAVATED ASSAUL...| 2|
| INVESTIGATIVE DET...| 2|
| ATTEMPTED ROBBERY...| 1|
| BURGLARY OF APART...| 1|
| STAY AWAY ORDER V...| 1|
+-----+

```

only showing top 20 rows

```

+-----+-----+
|first_record_date |latest_record_date |
+-----+-----+
|2015-05-13 18:00:00|2015-05-13 23:53:00|
+-----+-----+

```

```

+-----+-----+-----+
| Dates|      date_time|      month|
+-----+-----+-----+
|2015-05-13 23:53:00|2015-05-13 23:53:00|2015-01-01|
|2015-05-13 23:53:00|2015-05-13 23:53:00|2015-01-01|
|2015-05-13 23:33:00|2015-05-13 23:33:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:00:00|2015-05-13 23:00:00|2015-01-01|
+-----+-----+-----+

```

[PdDistrict]	[count]
NORTHERN	20
SOUTHERN	14
CENTRAL	13
TENDERLOIN	12
BAYVIEW	12
INKESTOE	9
TARAVAL	7
MISSION	6
PARK	5
RICHMOND	2

	Category	hour	count
0	LARCENY/THEFT	18	12
1	OTHER OFFENSES	18	4
2	NON-CRIMINAL	18	4
3	BURGLARY	18	2
4	WARRANTS	18	2
5	WEAPON LAWS	18	1
6	VEHICLE THEFT	19	3
7	WARRANTS	19	2
8	DRUNKENNESS	19	1
9	BURGLARY	19	1
10	ASSAULT	19	1
11	ROBBERY	19	1
12	OTHER OFFENSES	19	1
13	SUSPICIOUS OCC	19	2
14	NON-CRIMINAL	19	2
15	LARCENY/THEFT	19	13
16	OTHER OFFENSES	20	4
17	WEAPON LAWS	20	1
18	VANDALISM	20	3
19	VEHICLE THEFT	20	1
20	LARCENY/THEFT	20	8
21	NON-CRIMINAL	20	2
22	WARRANTS	21	1
23	ROBBERY	21	1
24	ASSAULT	21	1
25	OTHER OFFENSES	21	1
26	LARCENY/THEFT	21	5
27	NON-CRIMINAL	21	3
28	VANDALISM	22	1
29	OTHER OFFENSES	22	1
30	NON-CRIMINAL	22	2
31	ROBBERY	22	1
32	LARCENY/THEFT	22	2

35	OTHER OFFENSES	23	2
36	VEHICLE THEFT	23	2

[Dates]	[month]	[hour]	[week_day]	[year_month]	[month_day]	[date_number]
2015-05-13 23:53:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:53:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:33:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:33:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:30:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:30:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:30:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:30:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:30:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:00:00	2015-01-01	23	4	5	13	5245
2015-05-13 23:00:00	2015-01-01	23	4	5	13	5245
2015-05-13 22:50:00	2015-01-01	22	4	5	13	5245
2015-05-13 22:30:00	2015-01-01	22	4	5	13	5245
2015-05-13 22:30:00	2015-01-01	22	4	5	13	5245
2015-05-13 22:06:00	2015-01-01	22	4	5	13	5245
2015-05-13 22:00:00	2015-01-01	22	4	5	13	5245
2015-05-13 22:00:00	2015-01-01	22	4	5	13	5245
2015-05-13 22:00:00	2015-01-01	22	4	5	13	5245
2015-05-13 21:55:00	2015-01-01	21	4	5	13	5245
2015-05-13 21:40:00	2015-01-01	21	4	5	13	5245
2015-05-13 21:30:00	2015-01-01	21	4	5	13	5245

only showing top 20 rows

week_day	count
4	100

month_day	count
13	100

[address]	[count]
100 Block of JONE...	3
400 Block of HYDE ST	3
900 Block of BUSH ST	2
400 Block of BRAN...	2
JEFFERSON ST / LE...	2
1600 Block of WEB...	2
23RD ST / WISCONS...	2
0 Block of SANSOM...	2
1500 Block of FIL...	2
1900 Block of HIS...	2

only showing top 10 rows

```
[ 'LARCENY/THEFT', 'OTHER OFFENSES', 'NON-CRIMINAL', 'WARRANTS', 'VEHICLE THEFT', 'VANDALISM', 'ROBBERY', 'BURGLARY', 'ASSAULT', 'WEAPON LAWS' ]
[ '400 Block of HYDE ST', '100 Block of JONES ST', '1500 Block of FILLMORE ST', '1600 Block of WEBSTER ST', '1900 Block of MISSION ST', '23RD ST / WISCONSIN ST', '400 Block of BRANNAN ST', 'JEFFERSON ST / LEAVENWORTH ST', '0 Block of SANSOME ST', '900 Block of BUSH ST' ]
```

window size 200

Total time-6:05:21

```
pes2ug19cs013@pes2ug19cs013-VirtualBox:~/Desktop/proj$ python3 stream.py -f crime -b 200
Namespace(batch_size=200, endless=False, file='crime')
Waiting for connection on port 6100...
Connected to ('127.0.0.1', 33220)
Starting to stream crime dataset
0% | 10/4390 [00:50<6:05:21, 5.00s/it]
```

Results

```

pes2ugi9cs013@pes2ugi9cs013-VirtualBox:~/Desktop/proj$ /opt/spark/bin/spark-submit client.py 2>log.txt
+-----+-----+-----+-----+-----+-----+-----+-----+
| Dates | Category | Descript | DayOfWeek | PdDistrict | Resolution | Address | X | Y |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 2015-05-13 23:53:00 | WARRANTS | WARRANT ARREST | Wednesday | NORTHERN | ARREST, BOOKED | OAK ST / LAGUNA ST | -122.425891675136 | 37.7745985956747 |
| 2015-05-13 23:53:00 | OTHER OFFENSES | TRAFFIC VIOLATION... | Wednesday | NORTHERN | ARREST, BOOKED | OAK ST / LAGUNA ST | -122.425891675136 | 37.7745985956747 |
| 2015-05-13 23:33:00 | OTHER OFFENSES | TRAFFIC VIOLATION... | Wednesday | NORTHERN | ARREST, BOOKED | VANNES AV / GREE... | -122.42436302145 | 37.8004143219856 |
| 2015-05-13 23:30:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | NORTHERN | NONE | 1500 Block of LOM... | -122.426995326766 | 37.80087263276921 |
| 2015-05-13 23:30:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | PARK | NONE | 100 Block of BROD... | -122.438737622757 | 37.7715411720578 |
| 2015-05-13 23:30:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | INGLESIDE | NONE | 0 Block of TEDDY AV | -122.403252361212 | 37.713430704116 |
| 2015-05-13 23:30:00 | VEHICLE THEFT | STOLEN AUTOMOBILE | Wednesday | INGLESIDE | NONE | AVALON AV / PERU AV | -122.423326976668 | 37.7251380403778 |
| 2015-05-13 23:30:00 | VEHICLE THEFT | STOLEN AUTOMOBILE | Wednesday | BAYVIEW | NONE | KIRKWOOD AV / DON... | -122.371274317441 | 37.7275640719518 |
| 2015-05-13 23:00:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | RICHMOND | NONE | 600 Block of 47TH AV | -122.508194031117 | 37.7766012606812 |
| 2015-05-13 23:00:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | CENTRAL | NONE | JEFFERSON ST / LE... | -122.419087676747 | 37.8078015516515 |
| 2015-05-13 22:58:00 | LARCENY/THEFT | PETTY THEFT FROM ... | Wednesday | CENTRAL | NONE | JEFFERSON ST / LE... | -122.419087676747 | 37.8078015516515 |
| 2015-05-13 22:30:00 | OTHER OFFENSES | MISCELLANEOUS INV... | Wednesday | TARAVAL | NONE | 0 Block of ESCOLT... | -122.487983072777 | 37.73766665433271 |
| 2015-05-13 22:30:00 | VANDALISM | MALICIOUS MISCHIE... | Wednesday | TENDERLOIN | NONE | TURK ST / JONES ST | -122.412414263581 | 37.7830037964534 |
| 2015-05-13 22:00:00 | LARCENY/THEFT | GRAND THEFT FROM ... | Wednesday | NORTHERN | NONE | FILLMORE ST / GEA... | -122.432914603494 | 37.7843533426508 |
| 2015-05-13 22:00:00 | NON-CRIMINAL | FOUND PROPERTY | Wednesday | BAYVIEW | NONE | 200 Block of WILL... | -122.397744427103 | 37.7299346936044 |
| 2015-05-13 22:00:00 | NON-CRIMINAL | FOUND PROPERTY | Wednesday | BAYVIEW | NONE | 0 Block of MENDEL... | -122.383691503959 | 37.7431890419965 |
| 2015-05-13 22:00:00 | ROBBERY | ROBBERY, ARMED WI... | Wednesday | TENDERLOIN | NONE | EDDY ST / JONES ST | -122.412597377187 | 37.783932027727296 |
| 2015-05-13 21:55:00 | ASSAULT | AGGRAVATED ASSAUL... | Wednesday | INGLESIDE | NONE | GODEVS ST / MISSI... | -122.421681531572 | 37.7428222004845 |
| 2015-05-13 21:40:00 | OTHER OFFENSES | TRAFFIC VIOLATION | Wednesday | BAYVIEW | ARREST, BOOKED | MENDELL ST / HUDS... | -122.386400869953 | 37.738983491072 |
| 2015-05-13 21:30:00 | NON-CRIMINAL | FOUND PROPERTY | Wednesday | TENDERLOIN | NONE | 100 Block of JONE... | -122.412249767634 | 37.782556330202 |
+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
+-----+-----+-----+
| Category | count |
+-----+-----+-----+
| LARCENY/THEFT | 71 |
| OTHER OFFENSES | 24 |
| NON-CRIMINAL | 22 |
| WARRANTS | 15 |
| VANDALISM | 13 |
| VEHICLE THEFT | 12 |
| BURGLARY | 8 |
| ASSAULT | 6 |
| SUSPICIOUS OCC | 6 |
| ROBBERY | 6 |
| DRUG/NARCOTIC | 4 |
| FORGERY/COUNTERFE... | 3 |
| SECONDARY CODES | 2 |
| MISSING PERSON | 2 |
| TRESPASS | 2 |
| WEAPON LAWS | 2 |
| DRUNKENNESS | 1 |
| STOLEN PROPERTY | 1 |
+-----+-----+-----+

```

```

+-----+-----+-----+
| Descript | count |
+-----+-----+-----+
| GRAND THEFT FROM ... | 42 |
| AIDED CASE, MENTA... | 10 |
| PETTY THEFT FROM ... | 9 |
| WARRANT ARREST | 8 |
| STOLEN AUTOMOBILE | 7 |
| MALICIOUS MISCHIE... | 7 |
| MALICIOUS MISCHIE... | 6 |
| PETTY THEFT FROM ... | 5 |
| DRIVERS LICENSE, ... | 5 |
| GRAND THEFT FROM ... | 5 |
| INVESTIGATIVE DET... | 4 |
| FOUND PROPERTY | 4 |
| ENROUTE TO OUTSID... | 4 |
| LOST PROPERTY | 3 |
| TRAFFIC VIOLATION... | 3 |
| PETTY THEFT SHOPL... | 3 |
| TRAFFIC VIOLATION | 3 |
| MISCELLANEOUS INV... | 2 |
| STOLEN TRUCK | 2 |
| ENROUTE TO PAROLE... | 2 |
+-----+-----+-----+
only showing top 20 rows
+-----+-----+-----+
| first_record_date | latest_record_date |
+-----+-----+-----+
| 2015-05-13 14:03:00 | 2015-05-13 23:53:00 |
+-----+-----+-----+
+-----+-----+-----+
| Dates | date_time | month |
+-----+-----+-----+
| 2015-05-13 23:53:00 | 2015-05-13 23:53:00 | 2015-01-01 |
| 2015-05-13 23:53:00 | 2015-05-13 23:53:00 | 2015-01-01 |
| 2015-05-13 23:33:00 | 2015-05-13 23:33:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:30:00 | 2015-05-13 23:30:00 | 2015-01-01 |
| 2015-05-13 23:00:00 | 2015-05-13 23:00:00 | 2015-01-01 |
| 2015-05-13 23:00:00 | 2015-05-13 23:00:00 | 2015-01-01 |
| 2015-05-13 22:58:00 | 2015-05-13 22:58:00 | 2015-01-01 |
| 2015-05-13 22:30:00 | 2015-05-13 22:30:00 | 2015-01-01 |
| 2015-05-13 22:30:00 | 2015-05-13 22:30:00 | 2015-01-01 |
| 2015-05-13 22:00:00 | 2015-05-13 22:00:00 | 2015-01-01 |
+-----+-----+-----+

```

```

|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:30:00|2015-05-13 23:30:00|2015-01-01|
|2015-05-13 23:00:00|2015-05-13 23:00:00|2015-01-01|
|2015-05-13 23:00:00|2015-05-13 23:00:00|2015-01-01|
|2015-05-13 22:58:00|2015-05-13 22:58:00|2015-01-01|
|2015-05-13 22:30:00|2015-05-13 22:30:00|2015-01-01|
|2015-05-13 22:30:00|2015-05-13 22:30:00|2015-01-01|
|2015-05-13 22:06:00|2015-05-13 22:06:00|2015-01-01|
|2015-05-13 22:00:00|2015-05-13 22:00:00|2015-01-01|
|2015-05-13 22:00:00|2015-05-13 22:00:00|2015-01-01|
|2015-05-13 22:00:00|2015-05-13 22:00:00|2015-01-01|
|2015-05-13 21:55:00|2015-05-13 21:55:00|2015-01-01|
|2015-05-13 21:40:00|2015-05-13 21:40:00|2015-01-01|
|2015-05-13 21:30:00|2015-05-13 21:30:00|2015-01-01|
+-----+

```

only showing top 20 rows

```

+-----+
| Resolution | month | count |
+-----+
| PSYCHOPATHIC CASE | 2015-01-01 | 1 |
| JUVENILE BOOKED | 2015-01-01 | 2 |
| ARREST, CITED | 2015-01-01 | 3 |
| ARREST, BOOKED | 2015-01-01 | 35 |
| NONE | 2015-01-01 | 159 |
+-----+

```

```

      Resolution      month  count yearpd  arrest
0  PSYCHOPATHIC CASE  2015-01-01      1  2015   False
1  JUVENILE BOOKED    2015-01-01      2  2015   False
2  ARREST, CITED      2015-01-01      3  2015    True
3  ARREST, BOOKED     2015-01-01     35  2015    True
4      NONE           2015-01-01    159  2015   False

  hour  count
0     22      7
1     16     23
2     20     19
3     19     27
4     15     23
5     17     33
6     23     10
7     21     12
8     14     19
9     18     27

```

```

+-----+
| PdDistrict | count |
+-----+
| NORTHERN | 34 |
| SOUTHERN | 32 |
| CENTRAL | 26 |
| INGLESIDE | 24 |
| TENDERLOIN | 23 |
| BAYVIEW | 22 |
| MISSION | 12 |
| PARK | 11 |
| TARAVAL | 8 |
| RICHMOND | 8 |
+-----+

```

```

      Category  hour  count
0  DRUG/NARCOTIC    14      2
1  NON-CRIMINAL     14      2
2  VANDALISM        14      2
3  BURGLARY         14      2
4  WARRANTS         14      4
..  ...            ...    ...
73  LARCENY/THEFT    22      2
74  WARRANTS        23      1
75  LARCENY/THEFT    23      5
76  OTHER OFFENSES   23      2
77  VEHICLE THEFT    23      2

```

[78 rows x 3 columns]

```

+-----+
| Dates | month | hour | week_day | year_month | month_day | date_number |
+-----+
| 2015-05-13 23:53:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:53:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:33:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:30:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:30:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:30:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:30:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:30:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:00:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:00:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 23:00:00 | 2015-01-01 | 23 | 4 | 5 | 13 | 5245 |
| 2015-05-13 22:58:00 | 2015-01-01 | 22 | 4 | 5 | 13 | 5245 |
| 2015-05-13 22:30:00 | 2015-01-01 | 22 | 4 | 5 | 13 | 5245 |
| 2015-05-13 22:30:00 | 2015-01-01 | 22 | 4 | 5 | 13 | 5245 |
| 2015-05-13 22:06:00 | 2015-01-01 | 22 | 4 | 5 | 13 | 5245 |
| 2015-05-13 22:00:00 | 2015-01-01 | 22 | 4 | 5 | 13 | 5245 |
| 2015-05-13 22:00:00 | 2015-01-01 | 22 | 4 | 5 | 13 | 5245 |

```

```

[2015-05-13 22:30:00|2015-01-01|22 |4 |5 |13 |5245 |
[2015-05-13 22:00:00|2015-01-01|22 |4 |5 |13 |5245 |
[2015-05-13 22:00:00|2015-01-01|22 |4 |5 |13 |5245 |
[2015-05-13 22:00:00|2015-01-01|22 |4 |5 |13 |5245 |
[2015-05-13 22:00:00|2015-01-01|22 |4 |5 |13 |5245 |
[2015-05-13 21:55:00|2015-01-01|21 |4 |5 |13 |5245 |
[2015-05-13 21:40:00|2015-01-01|21 |4 |5 |13 |5245 |
[2015-05-13 21:30:00|2015-01-01|21 |4 |5 |13 |5245 |
+-----+
only showing top 20 rows
+-----+
week_day count
0 4 200
month_day count
0 13 200
+-----+
| address|count|
+-----+
[0 Block of WHITFI...| 4|
[1200 Block of MAR...| 3|
[100 Block of JONE...| 3|
[400 Block of HYDE ST| 3|
[ 6TH ST / NATOMA ST| 3|
[1900 Block of MIS...| 2|
[1500 Block of FIL...| 2|
[2300 ST / WISCONS...| 2|
[1600 Block of HAI...| 2|
[1600 Block of WEB...| 2|
+-----+
only showing top 10 rows
['LARCENY/THEFT', 'OTHER OFFENSES', 'NON-CRIMINAL', 'WARRANTS', 'VANDALISM', 'VEHICLE THEFT', 'BURGLARY', 'ASSAULT', 'SUSPICIOUS OCC', 'ROBBERY']
['0 Block of WHITFIELD CT', '400 Block of HYDE ST', '1200 Block of MARKET ST', '100 Block of JONES ST', '6TH ST / NATOMA ST', '300 Block of MCALLISTER ST', '1600 Block of HAIGHT ST', '1500 Block of FILLMORE ST', '1600 Block of WEBSTER ST', '1900 Block of MISSION ST']

```

Kafka producer and consumer

Start Zookeeper

1)cd /opt/kafka_2.13-3.1.0

2)sudo bin/zookeeper-server-start.sh
config/zookeeper.properties


```
[sudo] password for pes2zug19cs013:
2022-04-24 11:15:46,531 INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,543 WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,564 INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,564 INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,564 INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,565 INFO metricsProvider.className is org.apache.zookeeper.metrics.Tmpl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,570 INFO autopurge.snapshotCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
2022-04-24 11:15:46,580 INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
2022-04-24 11:15:46,588 INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
2022-04-24 11:15:46,589 WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
2022-04-24 11:15:46,607 INFO Log4j 1.2 jmx support found and enabled. (org.apache.zookeeper.jmx.ManagedUtil)
2022-04-24 11:15:46,622 INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,662 WARN config/zookeeper.properties is relative. Prepend ./ to indicate that you're sure! (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,663 INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,664 INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,664 INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,664 INFO metricsProvider.className is org.apache.zookeeper.metrics.Tmpl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
2022-04-24 11:15:46,664 INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
2022-04-24 11:15:46,739 INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.Tmpl.DefaultMetricsProvider@b2f6ed (org.apache.zookeeper.server.ServerMetrics)
2022-04-24 11:15:46,754 INFO zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
2022-04-24 11:15:46,830 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,830 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,830 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,830 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,830 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,831 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,832 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,832 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,833 INFO (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,839 INFO Server environment:zookeeper.version=3.6.3--64b1e4ad208761bc6b9f80dec2d69f2e3c866a, built on 04/08/2021 16:35 GMT (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,839 INFO Server environment:host.name=pes2zug19cs013-VirtualBox (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,840 INFO Server environment:java.version=11.0.4.1 (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,843 INFO Server environment:java.vendor=Ubuntu (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,843 INFO Server environment:java.home=/usr/lib/jvm/java-11-openjdk-amd64 (org.apache.zookeeper.server.ZooKeeperServer)
2022-04-24 11:15:46,843 INFO Server environment:java.class.path=/opt/kafka-2.13-3.1.0/bin/.:/lib/activation-1.1.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/aopalliance-repackaged-2.6.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/args4j-0.7.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/annotations-0.5.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/commons-cll-1.4.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/commons-lang3-3.8.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-api-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-basic-auth-extension-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-file-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-json-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-mirror-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-runtime-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/connect-transforms-3.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/hk2-engine-2.6.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/hk2-utils-2.6.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-annotations-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-core-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-databind-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-dataformat-csv-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-dataformat-jdk8-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-jaxrs-base-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-jaxrs-providers-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-module-jaxb-annotation-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-module-jakarta-xmlbind-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jackson-module-scala-2.12.3.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jetty-http-9.4.46.v20210219.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jetty-io-9.4.46.v20210219.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jetty-servlet-9.4.46.v20210219.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jetty-util-9.4.46.v20210219.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jetty-webapp-9.4.46.v20210219.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jetty-xml-9.4.46.v20210219.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/jsr305-3.0.2.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/kryo-4.0.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/log4j-1.2.17.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/log4j-api-2.17.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/maven-artifact-3.0.9.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/minlog-1.3.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-buffer-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-codec-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-common-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-handler-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-resolver-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-tcnative-2.0.43.Final-linux-x86_64.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-tcnative-classes-2.0.43.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-transport-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-transport-native-unix-common-4.1.60.Final.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/netty-transport-native-epoll-4.1.60.Final-linux-x86_64.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/osgi-4.2.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/packaging-0.1.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/range-3.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/reactor-netty-1.0.1.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/reload4j-1.2.20.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/slf4j-api-1.7.36.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/slf4j-log4j12-1.7.36.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/swagger-ui-3.10.2.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/swagger-annotations-2.1.14.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/swagger-models-2.1.14.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/swagger-parser-2.1.14.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/swagger-ui-swagger-project-v2.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/thrift-0.15.0.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-embed-el-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-embed-core-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-embed-war-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-jasper-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-jsp-api-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-servlet-api-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-util-scan-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-api-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-client-api-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-server-api-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-impl-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-parent-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-runtime-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-test-runtime-9.0.53.jar:/opt/kafka-2.13-3.1.0/bin/.:/lib/tomcat-websocket-transport-9.0.53.jar:/opt/k
```

Start kafka server

```
cd /opt/kafka 2.13-3.1.0
```

```
sudo JMX_PORT=8004 bin/kafka-server-start.sh
config/server.properties
```

```

pes2ugi9cs013@pes2ugi9cs013-VirtualBox: $ cd /opt/kafka_2.13-3.1.0
pes2ugi9cs013@pes2ugi9cs013-VirtualBox: /opt/kafka_2.13-3.1.0 $ sudo JMX_PORT=8084 bin/kafka-server-start.sh config/server.properties
2022-04-24 11:18:31,058] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration)
2022-04-24 11:18:32,259] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.XS09Util)
2022-04-24 11:18:32,618] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
2022-04-24 11:18:32,632] INFO starting (kafka.server.KafkaServer)
2022-04-24 11:18:32,634] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
2022-04-24 11:18:32,706] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181: (kafka.zookeeper.ZooKeeperClient)
2022-04-24 11:18:32,719] INFO Client environment:zookeeper.version=3.6.3--6401e4ad2007061bcb0f80dec2dd09f2e3c660a, built on 04/08/2021 16:35 GMT (org.apache.zookeeper.ZooKeeper)
2022-04-24 11:18:32,719] INFO Client environment:host.name=pes2ugi9cs013-VirtualBox (org.apache.zookeeper.ZooKeeper)
2022-04-24 11:18:32,719] INFO Client environment:java.version=11.0.14.1 (org.apache.zookeeper.ZooKeeper)
2022-04-24 11:18:32,720] INFO Client environment:java.vendor=Ubuntu (org.apache.zookeeper.ZooKeeper)
2022-04-24 11:18:32,720] INFO Client environment:java.home=/usr/lib/jvm/java-11-openjdk-amd64 (org.apache.zookeeper.ZooKeeper)
2022-04-24 11:18:32,720] INFO Client environment:java.class.path=/opt/kafka_2.13-3.1.0/bin/./libs/activation-1.1.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/aopalliance-repackaged-2.6.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/argparse4j-0.7.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/audience-annotations-0.5.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/commons-cl-1.4.jar:/opt/kafka_2.13-3.1.0/bin/./libs/commons-lang3-3.8.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-api-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-basic-auth-extension-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-flie-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-json-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-mirror-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-mirror-client-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-runtime-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/connect-transforms-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/hk2-api-2.6.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/hk2-locator-2.6.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/hk2-utils-2.6.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-annotations-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-core-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-databind-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-dataformats-csv-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-datatype-jdk8-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-jaxrs-base-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-jaxrs-json-provider-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-module-jaxb-annotations-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jackson-module-scala-2.12.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jakarta.activation-api-1.2.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jakarta.annotation-api-1.3.5.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jakarta.inject-2.6.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jakarta.validation-api-2.0.2.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jakarta.ws.rs-api-2.1.6.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jakarta.xml.bind-api-2.3.2.jar:/opt/kafka_2.13-3.1.0/bin/./libs/javassist-3.27.0-GA.jar:/opt/kafka_2.13-3.1.0/bin/./libs/javax.servlet-api-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/javax.ws.rs-api-2.1.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jaxb-api-2.3.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jersey-client-2.34.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jersey-common-2.34.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jersey-container-servlet-2.34.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jersey-container-servlet-core-2.34.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jersey-hk2-2.34.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jersey-server-2.34.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-client-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-continuation-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-http-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-io-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-security-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-server-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-servlet-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-servlets-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-util-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jetty-util-ajax-9.4.43.v20210629.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jline-3.12.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jopt-simple-5.0.4.jar:/opt/kafka_2.13-3.1.0/bin/./libs/jose4j-0.7.8.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-2.13-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-clients-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-log4j-appender-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-metadata-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-raft-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-server-common-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-shell-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-storage-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-storage-api-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-streams-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-streams-examples-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-streams-scala-2.13-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-streams-test-utils-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/kafka-tools-3.1.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/log4j-1.2.17.jar:/opt/kafka_2.13-3.1.0/bin/./libs/lz4-java-1.8.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/maven-artifact-3.8.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/metrics-core-2.2.0.jar:/opt/kafka_2.13-3.1.0/bin/./libs/metrics-core-4.1.12.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-buffer-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-codec-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-common-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-handler-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-resolver-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-transport-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-transport-native-epoll-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/netty-transport-native-unix-common-4.1.68.Final.jar:/opt/kafka_2.13-3.1.0/bin/./libs/osgi-resource-locator-1.0.3.jar:/opt/kafka_2.13-3.1.0/bin/./libs/paranamer-2.8.jar:/opt/kafka_2.13-3.1.0/bin/./libs/plexus-utils-3.2.1.jar:/opt/kafka_2.13-3.1.0/bin/./libs/reflections-0.9.12.jar

```

Start kafka configuration manager(CMAK)

cd /opt/CMAK/target/universal/cmak-3.0.0.6

sudo bin/cmak -Dconfig.file=conf/application.conf

-Dhttp.port=8080bin/cmak

-Dconfig.file=conf/application.conf -Dhttp.port=8080

```
pes2ugi5c-0t1gpes2ugi5c013-VirtualBox:/opt/CMAK/target/universal/cnak-3.0.0.$ sudo bin/cnak -Dconfig.file=conf/application.conf -Dhttp.port=8080bin/cnak -Dconfig.file=conf/application.conf -Dh
http.port=8080
2022-04-24 11:34:12,222 - [WARN] application - conf/application.conf: 12: play.crypto.secret is deprecated, use play.http.secret.key instead
2022-04-24 11:34:13,360 - [INFO] k.m.a.KafkaManagerActor - Starting curator...
2022-04-24 11:34:13,435 - [INFO] o.a.z.ZooKeeper - Client environment:zookeeper.version=3.5.7-f0fdd52973d37f9d9c80b8d99842dc2c7f600e, built on 02/10/2020 11:30 GMT
2022-04-24 11:34:13,435 - [INFO] o.a.z.ZooKeeper - Client environment:host.name=pes2ugi5c013-VirtualBox
2022-04-24 11:34:13,435 - [INFO] o.a.z.ZooKeeper - Client environment:java.version=11.0.14.1
2022-04-24 11:34:13,435 - [INFO] o.a.z.ZooKeeper - Client environment:java.vendor=Ubuntu
2022-04-24 11:34:13,435 - [INFO] o.a.z.ZooKeeper - Client environment:java.home=/usr/lib/jvm/java-11-openjdk-amd64
2022-04-24 11:34:13,435 - [INFO] o.a.z.ZooKeeper - Client environment:java.class.path=/opt/CMAK/target/universal/cnak-3.0.0.0/lib/.../conf:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/cnak-cnak-3
.0.0.0-sans-externalized.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.apache.commons.commons-lang3-3.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.apache.kafka.kafka-clients-2.4.1
.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/javax.inject.javax.inject-1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.fasterxml.jackson.core.jackson-core-2.10.0.jar:/opt/CMAK/target/u
niversal/cnak-3.0.0.0/lib/com.fasterxml.jackson.datatype.jakarta-2.8.11.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.scalaz.scalaz-core-2.12-7.2.27.jar:/opt/CMAK/target/un
iversal/cnak-3.0.0.0/lib/com.typesafe.play.play-2.12-2.6.21.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-codec-4.1.45.Final.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com
.yammer.metrics.metrics-core-2.2.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.fasterxml.jackson.datatype.jakarta-2.10.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org
.webjars.requirejs-2.3.5.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.apache.curator.curator-client-2.12.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.typesafe.play.play-functional
-2.12-2.6.12.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/commons-codec.commons-codec-1.11.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.beachape.enumeratum-necros-2.12-1.5.9.jar:/opt/C
MAK/target/universal/cnak-3.0.0.0/lib/javax.transaction.jta-1.1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-transport-4.1.45.Final.jar:/opt/CMAK/target/universal/cnak-3.0.0.0.
lib/org.webjars.popper.js-1.14.3.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.json4s.json4s-scalap-2.12-3.6.5.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.typesafe.play.play-server
-2.12-2.6.21.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.apache.kafka.connect-jackson-2.2.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.slf4j.jul-to-slf4j-1.7.25.jar:/opt/CMAK/targe
t/universal/cnak-3.0.0.0/lib/org.typelevel.macro-compat-2.12-1.1.1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.reactivestreams.reactive-streams-1.0.2.jar:/opt/CMAK/target/universal/cnak-
3.0.0.0/lib/com.typesafe.play.filters-helpers-2.12-2.6.21.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.fasterxml.jackson.core.jackson-annotations-2.10.0.jar:/opt/CMAK/target/universal/cna
k-3.0.0.0/lib/ch.qos.logback.logback-classic-1.2.3.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.typesafe.play.play-streams-2.12-2.6.21.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org
.rocksd.bjnl-5.15.10.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.webjars.webjars-play-2.12-2.6.3.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/javax.activation.javax.activation-a
pt-1.2.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.google.code.findbugs.jsr305-3.0.2.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.slf4j.jcl-over-slf4j-1.7.25.jar:/opt/CMAK/targe
t/universal/cnak-3.0.0.0/lib/joda-time.joda-time-2.9.9.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-transport-native-epoll-4.1.45.Final.jar:/opt/CMAK/target/universal/cnak-3.0
.0.0/lib/com.typesafe.play.build-link-2.6.21.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.github.luben.zstd-jni-1.4.3-1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.apache.curator.c
urator-recipes-2.12.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.scala-lang.modules.scala-parser-combinators-2.12-1.1.1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-bu
ffer-4.1.45.Final.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.lz4.lz4-java-1.6.0.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.json4s.json4s-scalaz-2.12-3.6.5.jar:/opt/CMAK/target
/universal/cnak-3.0.0.0/lib/com.google.guava.guava-23.6.1-jre.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-transport-native-unix-common-4.1.45.Final.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/li
b/org.scala-lang.scala-reflect-2.12.10.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.webjars.bootstrap-4.3.1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-common-4.1.45.Fi
nal.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.typesafe.akka.akka-slf4j-2.12-2.5.19.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.typesafe.play.play-json-2.12-2.6.12.jar:/opt/CMAK
/target/universal/cnak-3.0.0.0/lib/com.typesafe.play.play-netty-utils-2.6.21.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-resolver-4.1.45.Final.jar:/opt/CMAK/target/universal/c
nak-3.0.0.0/lib/com.google.j2objc.j2objc-annotations-1.1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.typesafe.akka.akka-http-core-2.12-10.0.15.jar:/opt/CMAK/target/universal/cnak-3.0.0.0
/lib/org.webjars.json-20121008-1.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.json4s.json4s-ast-2.12-3.6.5.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.beachape.enumeratum-2.12-1.5
.13.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/org.clapper.grizzled-slf4j-2.12-1.3.3.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/io.netty.netty-handler-4.1.45.Final.jar:/opt/CMAK/target
/universal/cnak-3.0.0.0/lib/org.scala-lang.modules.scala-collection-compat-2.12-2.1.2.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.google.errorprone.error-prone-annotations-2.1.3.jar:/opt
/CMAK/target/universal/cnak-3.0.0.0/lib/org.webjars.dustjs-linkedin-2.7.2.jar:/opt/CMAK/target/universal/cnak-3.0.0.0/lib/com.fasterxml.jackson.module.jackson-module-scala-2.12-2.10.0.jar:/opt/C
```

<http://127.0.0.1:8080>

← → ↺ 127.0.0.1:8080

➤ ★ ⚙ □ 🖱 ⋮

🔍 CMAK Cluster ▼

Clusters

Clusters

Active	Operations	Version
--------	------------	---------

Page 19 / 25

Clusters / Add Cluster

← Add Cluster

Cluster Name

crime

Cluster Zookeeper Hosts

127.0.0.1:2181

Kafka Version

3.1.0

☒ Enable JMX Polling (Set JMX_PORT env variable before starting kafka server)

JMX Auth Username

JMX Auth Password

☐ JMX with SSL

☐ Enable Logkafka

run producer

python3 kafka_server.py

tutorial.txt

×

kafka_server.py

```
1 from pathlib import Path
2 import producer_server
3
4
5 def run_kafka_server():
6     input_file = f"{Path(__file__).parents[0]}/police-department-calls-for-service.json"
7
8     producer = producer_server.ProducerServer(
9         input_file=input_file,
10         topic="org.spark.streaming",
11         bootstrap_servers="localhost:9092",
12         client_id="kafka-server"
13     )
14
15     return producer
16
17
18 def feed():
19     producer = run_kafka_server()
20     producer.generate_data()
21
22
23 if __name__ == "__main__":
24     feed()
```

Page 20 / 25

producer_server

```
tutorial.txt  kafka_server.py  producer_server.py
1 from kafka import KafkaPrbducer
2 import json
3 import time
4
5
6 class ProducerServer(KafkaProducer):
7
8     def __init__(self, input_file, topic, **kwargs):
9         super().__init__(**kwargs)
10        self.input_file = input_file
11        self.topic = topic
12
13    def generate_data(self):
14        with open(self.input_file) as f:
15            data = json.load(f)
16            for line in data:
17                message = self.dict_to_binary(line)
18                self.send(self.topic, message)
19                time.sleep(1)
20
21    def dict_to_binary(self, json_dict):
22        return json.dumps(json_dict).encode('utf-8')
```

run consumer

python3 consumer_server.py

```
tutorial.txt  consumer_server.py
1 from kafka import KafkaConsumer
2 import logging
3 import time
4 from pymongo import MongoClient
5 import json
6
7
8 logger = logging.getLogger(__name__)
9 client = MongoClient('mongodb://localhost:27017')
10 #collection = client.numtest.numtest
11
12 db = client["mydatabase"]
13 collection = db["data"]
14
15 with open('police-department-calls-for-service.json') as f:
16     file_data = json.load(f)
17
18
19 if isinstance(file_data, list):
20     Collection.insert_many(file_data)
21 else:
22     Collection.insert_one(file_data)
23
24
25 def run_consumer_server():
26     consumer = KafkaConsumer(
27         "org.spark.streaming",
28         bootstrap_servers="localhost:9092",
29         auto_offset_reset="earliest",
30         group_id="crime_consumer"
31     )
32
33     for message in consumer:
34         #collection.insert_many(data)
35         print(f"Message: {message.value.decode('utf-8')}")
36
37
38
39 if __name__ == "__main__":
40     run_consumer_server()
```

Python Tab Width: 8 Ln 14, Col 1 INS

```
pes2ug19cs013@pes2ug19cs013-VirtualBox:~/Desktop/kafka-crime2$ python3 consumer_server.py
Message: {"crime_id": "183653763", "original_crime_type_name": "Traffic Stop", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:57", "call_date_time": "2018-12-31T23:57:00.000", "disposition": "ADM", "address": "Geary Bl/divisadero St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653750", "original_crime_type_name": "Traf Violation Cite", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:54", "call_date_time": "2018-12-31T23:54:00.000", "disposition": "CIT", "address": "100 Blk Howard St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Geo-Override", "common_location": ""}
Message: {"crime_id": "183653746", "original_crime_type_name": "Passing Call", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:49", "call_date_time": "2018-12-31T23:49:00.000", "disposition": "HAN", "address": "3300 Block Of 20th Av", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Common Location", "common_location": "Stonestown Galleria, SF"}
Message: {"crime_id": "183653745", "original_crime_type_name": "Audible Alarm", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:47", "call_date_time": "2018-12-31T23:47:00.000", "disposition": "PAS", "address": "1900 Block Of 18th Av", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Premise Address", "common_location": ""}
Message: {"crime_id": "183653737", "original_crime_type_name": "Traffic Stop", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:46", "call_date_time": "2018-12-31T23:46:00.000", "disposition": "CIT", "address": "Sansome St/chestnut St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653719", "original_crime_type_name": "Passing Call", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:38", "call_date_time": "2018-12-31T23:38:00.000", "disposition": "ND", "address": "Haight St/killmore St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653722", "original_crime_type_name": "Traffic Stop", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:30", "call_date_time": "2018-12-31T23:30:00.000", "disposition": "CIT", "address": "Geary Bl/parker Av", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653711", "original_crime_type_name": "Traffic Stop", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:36", "call_date_time": "2018-12-31T23:36:00.000", "disposition": "HAN", "address": "Battery St/union St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653706", "original_crime_type_name": "Passing Call", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:34", "call_date_time": "2018-12-31T23:34:00.000", "disposition": "Not recorded", "address": "1500 Block Of Haight St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Common Location", "common_location": "Haight St Corridor, SF"}
Message: {"crime_id": "183653717", "original_crime_type_name": "Fight No Weapon", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:33", "call_date_time": "2018-12-31T23:33:00.000", "disposition": "GOA", "address": "5th St/minna St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653702", "original_crime_type_name": "Traffic Stop", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:33", "call_date_time": "2018-12-31T23:33:00.000", "disposition": "HAN", "address": "Front St/green St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653703", "original_crime_type_name": "Passing Call", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:33", "call_date_time": "2018-12-31T23:33:00.000", "disposition": "HAN", "address": "3100 Block Of Mission St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Common Location", "common_location": "Roccapulco, SF"}
Message: {"crime_id": "183653700", "original_crime_type_name": "Passing Call", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:33", "call_date_time": "2018-12-31T23:33:00.000", "disposition": "HAN", "address": "Funston Av/cabrillo St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
Message: {"crime_id": "183653696", "original_crime_type_name": "Att", "report_date": "2018-12-31T00:00:00.000", "call_date": "2018-12-31T00:00:00.000", "offense_date": "2018-12-31T00:00:00.000", "call_time": "23:30", "call_date_time": "2018-12-31T23:30:00.000", "disposition": "REP", "address": "4000 Block Of 18th St", "city": "San Francisco", "state": "CA", "agency_id": "1", "address_type": "Intersection", "common_location": ""}
```

In CMAK topic org.spark.streaming is created and data is produced and consumed

+ org.spark.streaming

Topic Summary

Replication	1
Number of Partitions	1
Sum of partition offsets	544
Total number of Brokers	1
Number of Brokers for Topic	1
Preferred Replicas %	100
Brokers Skewed %	0
Brokers Leader Skewed %	0
Brokers Spread %	100
Under-replicated %	0

Metrics

Rate	Mean	1 min	5 min	15 min
Messages in /sec	1.00	1.00	1.03	1.11
Bytes in /sec	0.5k	0.5k	0.5k	0.5k

Operations

Delete Topic	Reassign Partitions	Generate Partition Assignments
Add Partitions	Update Config	Manual Partition Assignments

Partitions by Broker

Broker	# of Partitions	# as Leader	Partitions	Skewed?	Leader Skewed?
0	1	1	(0)	false	false

Consumers consuming from this topic

crime_consumer	KF
----------------	----

Number of Brokers for Topic	1
Preferred Replicas %	100
Brokers Skewed %	0
Brokers Leader Skewed %	0
Brokers Spread %	100
Under-replicated %	0

Metrics

Rate	Mean	1 min	5 min	15 min
Messages in /sec	1.00	1.00	1.03	1.11
Bytes in /sec	0.5k	0.5k	0.5k	0.5k
Bytes out /sec	0.5k	0.5k	0.5k	1.7k
Bytes rejected /sec	0.00	0.00	0.00	0.00
Failed fetch request /sec	0.00	0.00	0.00	0.00
Failed produce request /sec	0.00	0.00	0.00	0.00

Partition Information

Partition	Latest Offset	Leader	Replicas	In Sync Replicas	Preferred Leader?	Under Replicated?
0	544	0	(0)	(0)	true	false

Partitions by Broker

Broker	# of Partitions	# as Leader	Partitions	Skewed?	Leader Skewed?
0	1	1	(0)	false	false

Consumers consuming from this topic

crime_consumer	KF
----------------	----

← → ↺ ⓘ 127.0.0.1:8080/clusters/crime/consumers/crime_consumer/topic/org.spark.streaming/type/KF 🔍 ➤ ☆ 🌐 👤

🔴 CMAK **crime** Cluster ▾ Brokers Topic ▾ Preferred Replica Election Schedule Leader Election Reassign Partitions Consumers

Clusters / crime / Consumers / crime_consumer / org.spark.streaming

+ crime_consumer / org.spark.streaming

Topic Summary

Total Lag	1
% of Partitions assigned to a consumer instance	100

org.spark.streaming

Partition	LogSize	Consumer Offset	Lag	Consumer Instance Owner
0	971	970	1	kafka-python-2.0.2-ded6f6d2-e1b5-4a97-9228-c41387eb4dd0/127.0.0.1

Database-records added to mongodb

MongoDB Compass - localhost:27017/mydatabase.data

Connect View Collection Help

localhost:27017

7 DBS 7 COLLECTIONS

HOST localhost:27017

CLUSTER Standalone

EDITION MongoDB 3.6.8 Community

My Queries

Databases

Filter your data

- admin
- config
- database
- local
- mydatabase
- data
- numtest

Documents mydatabase.data

mydatabase.data

DOCUMENTS 400.0k STORAGE SIZE 45.1MB AVG. SIZE 445B INDEXES 1 TOTAL SIZE 3.4MB AVG. SIZE 3.4MB

Documents Aggregations Schema Explain Plan Indexes Validation

Filter { field: 'value' }

ADD DATA VIEW

Displaying documents 1 - 20 of 399998

```
{
  "_id": ObjectId("62664c59e2ad0dc9d9f1adf5"),
  "crime_id": "183653763",
  "original_crime_type_name": "Traffic Stop",
  "report_date": "2018-12-31T00:00:00.000",
  "call_date": "2018-12-31T00:00:00.000",
  "offense_date": "2018-12-31T00:00:00.000",
  "call_time": "23:57",
  "call_date_time": "2018-12-31T23:57:00.000",
  "disposition": "ADM",
  "address": "Geary Bl/divisadero St",
  "city": "San Francisco",
  "state": "CA",
  "agency_id": "1",
  "address_type": "Intersection",
  "common_location": ""
}
```

```
{
  "_id": ObjectId("62664c59e2ad0dc9d9f1adf6"),
  "crime_id": "183653756",
  "original_crime_type_name": "Traf Violation Cite",
  "report_date": "2018-12-31T00:00:00.000",
  "call_date": "2018-12-31T00:00:00.000",
  "offense_date": "2018-12-31T00:00:00.000",
  "call_time": "23:54",
  "call_date_time": "2018-12-31T23:54:00.000",
  "disposition": "CIT",
  "address": "100 Blk Howard St",
  "city": "San Francisco",
  "state": "CA",
  "agency_id": "1",
  "address_type": "Geo-Override",
  "common_location": ""
}
```


Conclusion

Data was streamed using spark streaming and spark queries were performed.

Data was also produced and consumed using kafka.

References

<https://spark.apache.org/>

<https://kafka.apache.org/documentation/#gettingStarted>