Final Project

Fraud Detection

CPE374 High Performance Computing & Parallel Technology

57070501024 Piyaphon Trangjirasathian

57070501038 Warat Kaweepornpoj

57070501078 Satit Udompanich

- Ingests
- ETL
- Develop Model
- Detect and Alert

```
import json
import os

text_file = open("json_log_huge.csv", "w")
for filename in os.listdir('/home/training/dataset/huge_json_data/'):
    with open('/home/training/dataset/huge_json_data/' + filename) as fp:
        for line in fp:
            data = json.loads(line)
            text_file.write(str(data['id']) + "," + str(data['timestamp']) + "," + str(data['channel']) +
             "," + str(data['userid']) + "," + str(data['action']) + "," + str(data['amount']) +
             "," + str(data['location']) + "\n")

text_file.close()
```

- Ingests
- ETL
- Develop Model
- Detect and Alert

- Ingests
- ETL
- **Develop Model**
- Detect and Alert

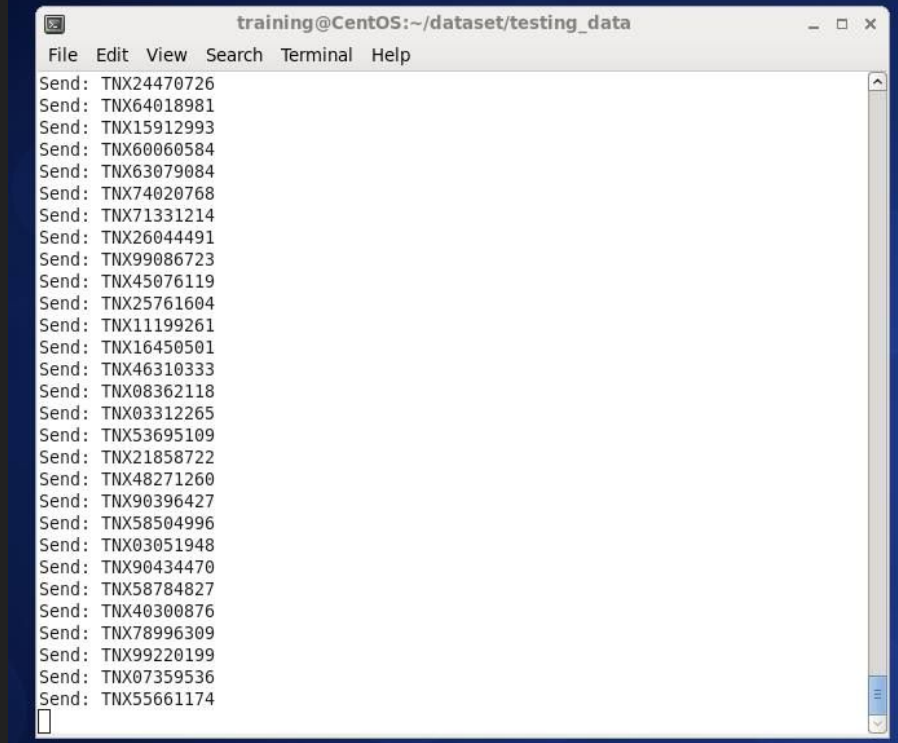
```
#!/usr/bin/python

from pyspark import SparkContext, SparkConf
import math
conf = SparkConf().setAppName("wordCount").setMaster("local[1]")

sc = SparkContext(conf=conf)

wordSet = sc.textFile("hdfs://localhost/user/training/csv/*.csv")
wordCount = wordSet.map(lambda line: line.split(",")).map(lambda word:(word[3],float(word[5])))
Num = wordCount.countByKey()
Sum = wordCount.reduceByKey(lambda a,b:a+b)
mean = Sum.map(lambda word: (word[0],(word[1]/Num[word[0]])))
meansq = Sum.map(lambda word: (word[0],(word[1]/Num[word[0]])**2))
Sumsq = wordCount.map(lambda word: (word[0],word[1]*word[1])).reduceByKey(lambda a,b:a+b).map(lambda word: (word[0],word[1]/Num[word[0]]))
SD = Sumsq.join(meansq).map(lambda word: (word[0], math.sqrt(word[1][0]-word[1][1])))
FraudCase = mean.join(SD).map(lambda word : (word[0],word[1][0]+(2*word[1][1]))).sortByKey(True)

# sd = sqrt(sumsq/n - meansq)

FraudCase.collect()
FraudCase.saveAsTextFile("hdfs://localhost/user/training/sparks1out")
```

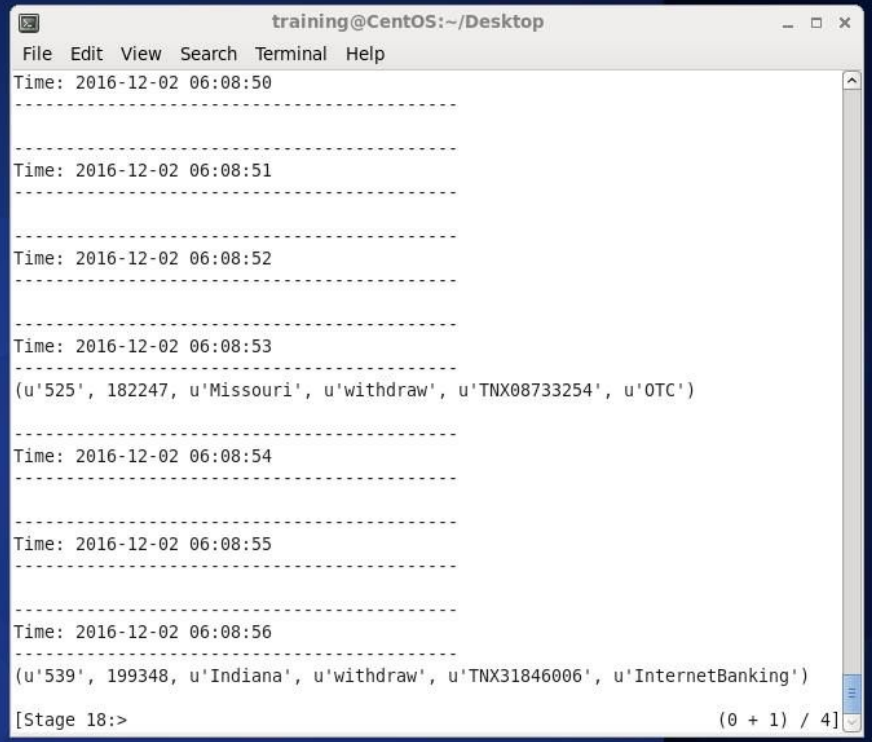
- Ingests
- ETL
- Develop Model
- Detect and Alert

```
#!/usr/bin/python

from pyspark import SparkContext, SparkConf
from pyspark.streaming import StreamingContext
import json
import pprint
import re

import smtplib
from email.mime.text import MIMEText
from email.mime.application import MIMEApplication
from email.mime.multipart import MIMEMultipart
from smtplib import SMTP
import sys
```

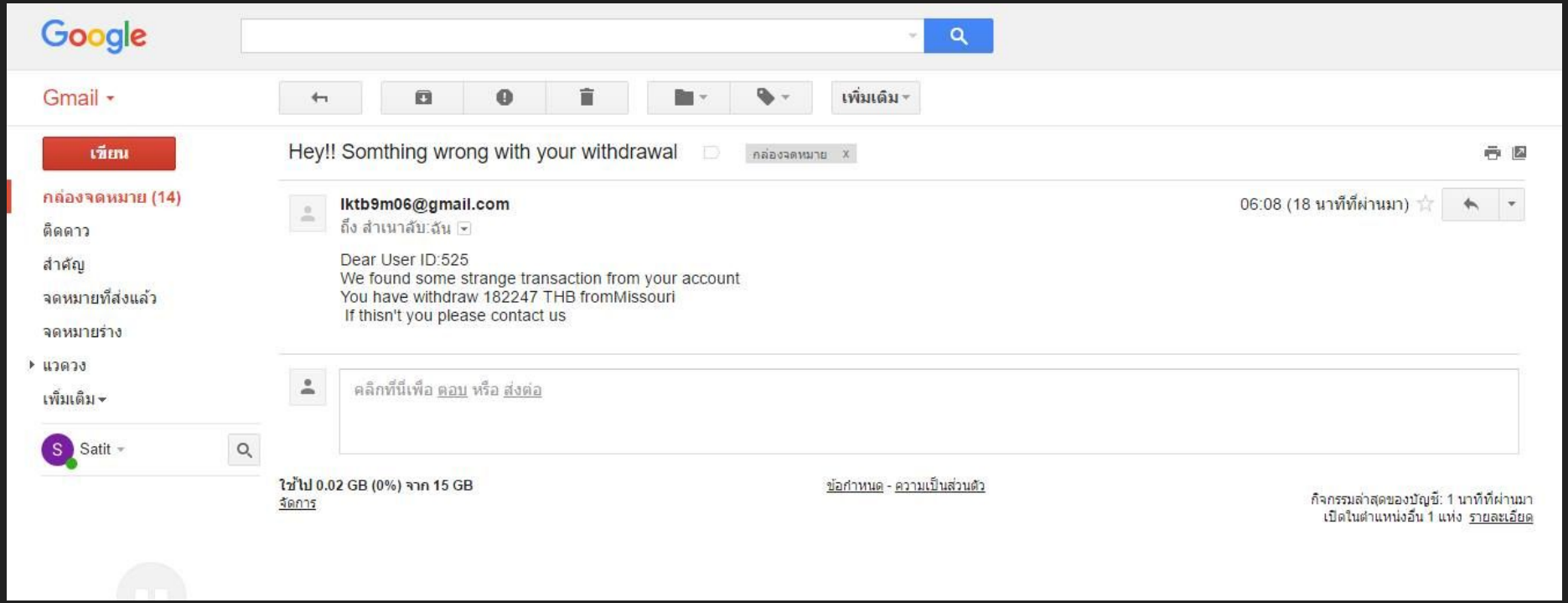
```python
def alert(x):
    #email alert function
    recipients = ['lktb9m06@gmail.com']
    emaillist = [elem.strip().split(',') for elem in recipients]
    msg = MIMEMultipart()
    msg['Subject'] = "Hey!! Somthing wrong with your withdrawal"
    msg['From'] = 'lktb9m06@gmail.com'
    msg['Reply-to'] = 'lktb9m06@gmail.com'
    msg.preamble = 'Multipart massage.\n'
    part = MIMEText("We found some strange transaction from your account\n\nuserid: " + str(x[0]) + "\namount: " +
    str(x[1]) + " THB\n location: " + str(x[2]) + "\n action: " + str(x[3]) + "\nid: " + str(x[4]) + "\nchannel: " + str(x[5]))
    msg.attach(part)

    server = smtplib.SMTP("smtp.gmail.com:587")
    server.ehlo()
    server.starttls()
    server.login("lktb9m06@gmail.com", "028184855")

    server.sendmail(msg['From'], emaillist, msg.as_string())

    return x
```

```
conf = SparkConf().setAppName("wordCount").setMaster("local[1]")
sc = SparkContext(conf=conf)
fraudFile = sc.textFile("hdfs://localhost/user/training/sparks1out/*")
fraudData = fraudFile.map(lambda x: re.sub("[\(u' )]","", x)).map(lambda x: x.split(",")).map(lambda x: (x[0],x[1])).collectAsMap()
sc.stop()

sc = SparkContext("local[2]", "NetworkWordCount")
ssc = StreamingContext(sc, 1)
lines = ssc.socketTextStream("localhost", 3222)


counts = lines.map(lambda x: json.loads(x)).map(lambda x: (x["userid"],x["amount"],x["location"],x["action"],x["id"],
x["channel"])).filter(lambda x: x[1] >= float(fraudData[x[0]])).map(lambda x: alert(x))
counts.pprint()

ssc.start()
ssc.awaitTermination()
```


training@CentOS:~/dataset/testing_data
File Edit View Search Terminal Help
Send: TNX24470726
Send: TNX64018981
Send: TNX15912993
Send: TNX60060584
Send: TNX63079084
Send: TNX74020768
Send: TNX71331214
Send: TNX26044491
Send: TNX99086723
Send: TNX45076119
Send: TNX25761604
Send: TNX11199261
Send: TNX16450501
Send: TNX46310333
Send: TNX08362118
Send: TNX03312265
Send: TNX53695109
Send: TNX21858722
Send: TNX48271260
Send: TNX90396427
Send: TNX58504996
Send: TNX03051948
Send: TNX90434470
Send: TNX58784827
Send: TNX40300876
Send: TNX78996309
Send: TNX99220199
Send: TNX07359536
Send: TNX55661174


training@CentOS:~/Desktop
File Edit View Search Terminal Help
Time: 2016-12-02 06:08:50
Time: 2016-12-02 06:08:51
Time: 2016-12-02 06:08:52
Time: 2016-12-02 06:08:53
(u'525', 182247, u'Missouri', u'withdraw', u'TNX08733254', u'OTC')
Time: 2016-12-02 06:08:54
Time: 2016-12-02 06:08:55
Time: 2016-12-02 06:08:56
(u'539', 199348, u'Indiana', u'withdraw', u'TNX31846006', u'InternetBanking')
[Stage 18:>
(0 + 1) / 4]

Result


Google
Gmail
เพิ่มเติม
เขียน
Hey!! Somthing wrong with your withdrawal
กล่องจดหมาย x
กล่องจดหมาย (14)
ติดดาว
สำคัญ
จดหมายที่ส่งแล้ว
จดหมายร่าง
แวดวง
เพิ่มเติม
Satit
lktb9m06@gmail.com
ถึง สำเนาลับ:ฉัน
06:08 (18 นาทีที่ผ่านมา)
Dear User ID:525
We found some strange transaction from your account
You have withdraw 182247 THB fromMissouri
If thisn't you please contact us
คลิกที่นี่เพื่อ ตอบ หรือ ส่งต่อ
ใช้ไป 0.02 GB (0%) จาก 15 GB
จัดการ
ข้อกำหนด - ความเป็นส่วนตัว
กิจกรรมล่าสุดของบัญชี: 1 นาทีที่ผ่านมา
เปิดในตำแหน่งอื่น 1 แห่ง รายละเอียด

Demo