

# Crime Analytics and Alert System

Anshul Patni, Nikhila Galala, Sanjay Raghavendra Karnati, Varun Jindal

*Software Engineering, San Jose State University  
One Washington Square, San Jose, CA, USA 95192*

{anshul.patni, nikhila.galala, sanjayraghavendra.karnati, varun.jindal}@sjsu.edu

**Abstract**—This paper presents a crime analytics and alert system which sends alerts when a violent crime has been reported in an area and analyzes the cause of these crimes in that region. The violent crime rate is on the rise throughout the United States. The problem of crime can occur any time anywhere. We need to actively think about our privacy and security. As the crime rates are increasing and violent crimes are becoming more sophisticated, protecting ourselves and our loved ones should be our priority. The crime alert system is an application which uses streaming technology to pick the violent crime from a control center database and send these crime alerts to the people residing in that area. It allows users to register to an alert system to receive alerts in real time. It also collects the demographic data of the people that area to interpret and analyze the reason or possibility of a crime in that region. This facilitates the police force to analyze and incorporate police officers focused to avoid certain crime, thus making an area safer to live.

**Keywords**—Demographic data, Spark, HDFS, Kafka, Hive.

## I. INTRODUCTION

The issue of security against heinous crimes has proved to be of utmost importance recently, more so than ever. This gave an interesting idea of making a crime alert system which would facilitate people to be aware about their surrounding all the time and help them stay safe and secure. This is not just an idea, but an area with great scope for making an impact, for overcoming the challenge of security and understanding the reasons visually. A crime alert system with an ability to analyze the causes of a crime in an area and its dependency on certain factors could provide us with an extremely useful insights and opportunities.

A crime alert system is an application that leverages data analytics on crime data and demographic data to provide solutions to the end users. The analytics is done on the data to find the possible causes of a certain crime happening a lot more in some areas than in others. This is achieved by visualizing the demographic factors like average income, average rental price, number of dropouts, number of homeless price, and other factors in a region and comparing it with crime data of that area to find possible relation or dependency of that factor contributing to the cause of that crime.

The correlation between certain demographic factors, economy of an area, among other factors and their analysis could be extremely helpful to the citizens by sending out crime alerts in real-time, making people aware about the crime situation in their area, thus helping them to avoid confronting a disastrous situation. This can also be used by the government officials to curb the crime rate by incorporating police force towards critical areas responsible for these crimes.

## II. NEED FOR CRIME ANALYTICS SYSTEM

### A. Security

In today's world, one needs to be proactive about their security and privacy. As the crime rates are increasing and the heinous are becoming more sophisticated, protecting ourselves and people around us should be our top priority. No one knows, what tragedy might occur at any point in their lives. We cannot predict what the future holds for us, but the one thing we need to do is to put up safeguards to prevent or at least minimize its effect.

The crime alert system can be used by citizens by subscribing to a city's crime alerts via mails or text messages or both. This will spread the awareness about a crime in the neighborhood in real-time. These alerts would help the citizens or the users using this application to avoid crime scenes and crime prone areas which will reduce the casualties due to crime and may help in reducing the violent crime rates in a region.

### B. Curbing the crime rates with the use of data analytics

Understanding the cause of a certain type of crime, which might be happening a lot more in an area than others, can be complicated. Not only does the nature of crime keeps changing, but also the mindset of people committing these crimes.

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also, the modern technologies and hi-tech methods help criminals in achieving their misdeeds, According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence.

A major challenge faced by most of the law enforcement and intelligence organizations is efficiently and accurately analyzing the growing volumes of crime related data. The vast geographical diversity and the complexity of crime patterns have made the analyzing and recording of crime data more difficult.

Data mining is a powerful tool that can be used effectively for analyzing large databases and deriving important analytical results. This paper presents an intelligent crime analysis system which is designed to overcome the above-mentioned problems.

### III. COMPONENTS OF CRIME ANALYTICS AND ALERT SYSTEM

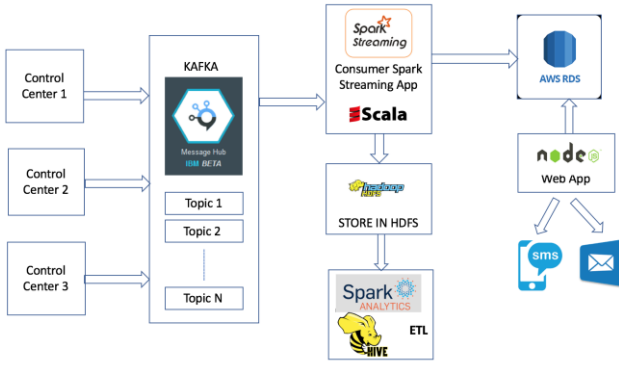


Fig 1. Architecture of the crime analytics and alert system

#### A. Data Ingestion Model

Data is ingested from control centers where all the crimes are reported for each city. We can consider this as a Police Central Crime reporting center where all the cops report incidents in real-time. These crime control centers will then ingest the specific crime related messages to KAFKA topics in real-time.

Kafka is used as broker between the crime control centers knows as producers and Spark based real time data ingestion application known as consumer. The reason to choose Kafka for this application is as follows

1. Kafka is real-time messaging queue which can scale for multiple thousands of messages read and writes per second.
2. Kafka has inbuilt mechanism for data replication and failure scenarios.
3. Kafka is designed for highly distributed cluster based application and with deep integration with Hadoop, Spark and Hadoop Distributed File System based distributed systems.
4. Kafka cluster (using IBM Message Hub) is configured with five brokers to handle any node failures and topics are created with partitions to allow higher read and write throughput. Control centers write data messages into Kafka using python using Kafka-confluent api's.

#### B. Spark Streaming

Apache Spark provides interface for reading and writing data from Kafka in real-time using Apache Spark Streaming. We use spark streaming in this application to read data from Kafka topics in real-time and write the data to HDFS for each data streams. Streams are read from Kafka in intervals of every 30 seconds, i.e. messages received in last 30 seconds are read and processed into HDFS.

Apache Spark Streaming provides data in the form of data streams also known as Discretized Stream or DStream which is basic level of data abstraction provided by spark streaming application.



Fig 2. Data abstraction using spark streaming

```
val conf = new SparkConf()
val ssc = new StreamingContext(conf, Seconds(30))
val sc = ssc.sparkContext
val messages = KafkaUtils.createDirectStream[String, String](
  ssc,
  LocationStrategies.PreferConsistent,
  ConsumerStrategies.Subscribe[String, String](
    topic.split(",").toSet, kafkaProperties))
```

Any data operations on Data Streams are translated into Apache Spark RDD operations. In our applications, we use map operations to perform filter and store in HDFS.

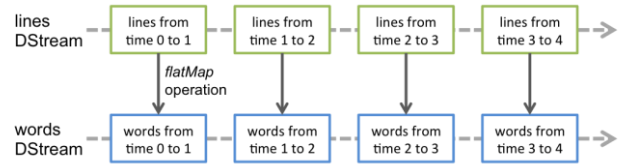


Fig 3. Data stream transformations

```
val kv = messages.map(a => (a.key(), a.value()))
// Filter RDD by city
val kv_table1 = kv.filter(_._1 == "chicago")
// Write DStreams to HDFS
kv_table1.foreachRDD(a => {
  if(a.count() > 0){
    val ts = System.nanoTime()
    a.map(_._2).saveAsTextFile("hdfs:///data/stream-out/crime_data/chicago/ts=" + System.nanoTime())
  } })
```

#### C. Data Transformations and Analytics

Data Transformations are performed on data residing in HDFS written by spark streaming application. Apache Spark provides abstractions to read data from hive (data warehouse for HDFS) as RDD's on top of which we can perform ETL (Extract Transform and Load) based operations to join with other datasets and standardize data across many datasets.

```
val sc =
  SparkSession.builder().enableHiveSupport().getOrCreate()
val df = sc.sql("select Description, year, COUNT(*) as
  crime_count from crime_etl.crime_etl_temp_chicago GROUP
  BY DESCRIPTION, year")

df.withColumn("City", lit("Chicago")).write.mode("overwrite")
  .saveAsTable("crime_etl.chicago_crime_type")
```

Apache Hive is an open source data warehouse implementation for data residing in HDFS. We can connect to hive using JDBC connection which makes it's a best suitable candidate to build real-time dashboards using third party applications like. Hive provides SQL like interface which translates all the queries into map-reduce operations making querying HDFS data easier. Hive can also be accessed from spark application as shown in above code snippet.

We have made a NodeJS application where users can register and login into the Web Application and subscribe for their interested crime alerts and the interested city. This would register the users to the corresponding AWS SNS topic of their city to send alerts. Below is the screen of the homepage of the NodeJS web application.

A screenshot of a web application for a 'Crime Alert System'. The page has a solid red background. At the top, there is a navigation bar with white text links: 'Home', 'Login', 'Register', 'Recent Alerts', and 'Crime Analysis'. The main heading is 'Crime Alert System' in a large, white, serif font. Below the heading is a white text prompt: 'Click here to see the latest Alerts on crime near you'. Underneath this prompt is a black rectangular button with the white text 'See Alerts'. At the bottom of the page, there is a white text paragraph: 'Are you afraid of going out? It is safe to go out around this time if your looking answers for these question then you have come to right place. Here in crime alert system we keep track of every crime near you and keep you updated.' To the right of this paragraph is a small, rectangular image of a book cover titled 'CRIME ALERT' in large, bold, white letters. The cover art depicts a dark, stormy sky with bright, glowing light sources, possibly representing a crime scene or a dramatic event.

City	Crime Count	Average Rent	Population
Chicago	27000	500	2700000
Los Angeles	22000	1500	3000000
San Francisco	15000	1500	800000

Apart from viewing the analytics user can register, login and subscribe to crime in a city of their interested type of crimes For example. Murder, theft, kidnap etc. The below screenshot shows user registration for a particular kind of crime. When the users registers he is subscribed to the particular AWS SNS topic to receive alerts.

Crime alert

Home

Profile Details

Emergency Contact

Logout

Alert PopUp

Register to get alerts.

☐ Email ☐ SMS

☐ SFO ☐

☐ assault ☐ theft ☐ murder ☐ kidnapping

Register


About Me

Name	City	Username	Email	Contact Number
Nikhil	themont	nikhilg23	nikhila.ganai@sjcu.edu	5105797999

User can subscribe to both email and SMS alerts, an email confirmation is sent to the user for the subscription to the specified email-id of the user during registration.

AWS Notification - Subscription Confirmation ✕ inbox x

---

 **SFO\_ALERTS** no-reply@sns.amazonaws.com via amazonses.com  
to me ▾

You have chosen to subscribe to the topic:  
**arn:aws:sns:us-west-2:390872976456:SFO\_ALERTS**

To confirm this subscription, click or visit the link below (If this was in error no action is necessary):  
[Confirm subscription](#)

Please do not reply directly to this email. If you wish to remove yourself from receiving all future SNS subscription confirmation requests please send an email to [sns-opt-out](#).

User can confirm the subscription and henceforth he would be receiving mails for the kind of crimes he has subscribed for. We are filtering the crime records after receiving the data from Kafka with the keyword such as theft, kidnap, murder and writing those records in AWS RDS. When admin logs in, he can view the all latest records. These records are displayed from the AWS RDS. If any of the data is un-notified then the admin can notify by clicking the notify button.

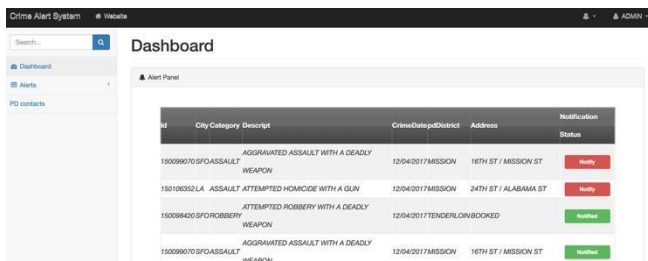


Fig 9. Crime Alert application admin dashboard

The above screenshot shows the admin dashboard where he has right to notify the alerts. On click of Notification the users receive the email and SMS alerts. The sample email and SMS notifications are shown below.

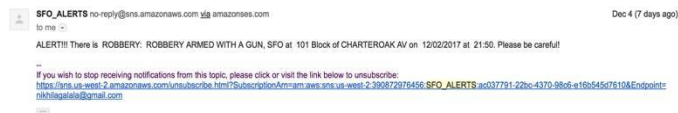


Fig 10. Email alert for a crime

The above screenshot shows the email notification. User also receive SMS alerts as shown in the figure below.

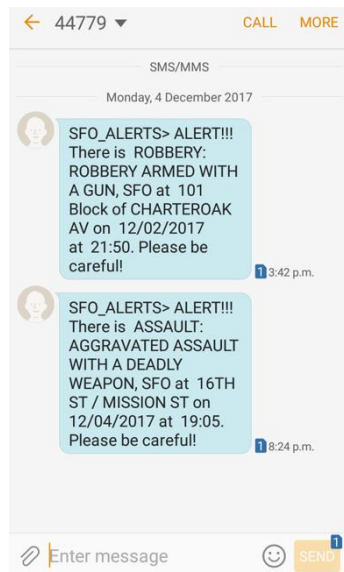


Fig 11. SMS alert for a crime

## E. Visualization

Apart from displaying the graphs in NodeJS web application we have also analyzed the statistics in Tableau. Below are the screenshots of the visuals.

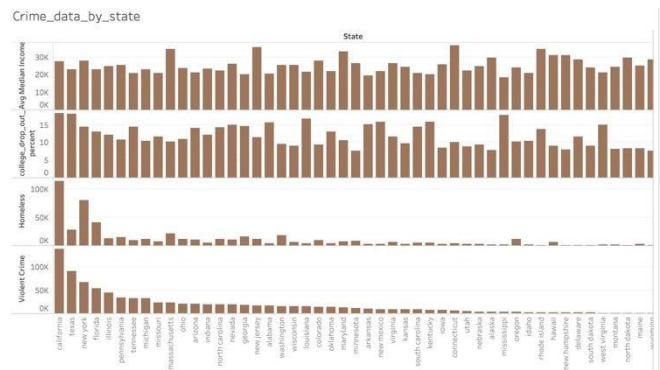


Fig 12. Visualizing crime data by states

Visualizing the crime data with the demographic data by states, the probable causes of the crimes and factors leading to them could be found. Looking at the data visualization, the violent crimes in states like California and Texas comes out since the number of college dropouts in these states are high. Another factor that might be leading to these crimes is the number of homeless people in a state.

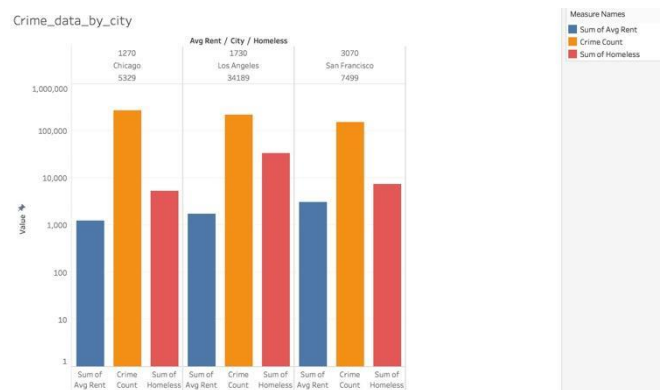


Fig 13. Visualizing crime data by cities

However, looking at the granular level and analyzing the city-wise crime data, it can be again be seen that somehow the number of violent crimes go high as the number of homeless people is increased. One interesting analysis that comes out from this visualization is that the number of criminal activities in a city is not exactly proportional to the average rental prices in that city. This shows that there are few factors which affect the number of crimes more than other factors.

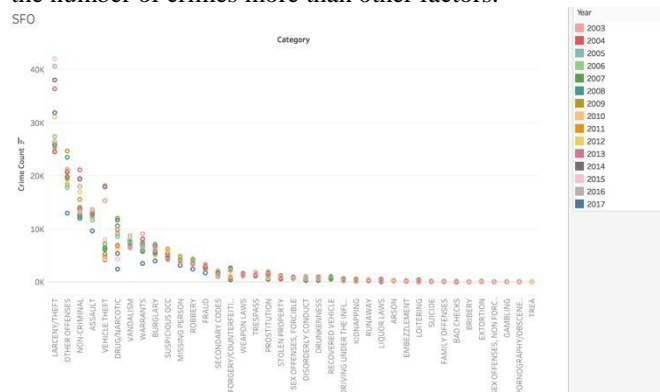


Fig 14. Visualizing crime data according to the nature of crime

Visualizing the crime data for a city by year for last few years shows us the crime that are on the increase and others which may be under control now. This data can be used to incorporate law enforcements in the areas which have been on the increase to curb the violent crimes in a region.

#### F. Implementation on Cloud

IBM Bluemix Message Hub is used for messaging queue (Kafka) which is deployed on IBM Cloud. Kafka is configured with 5 Broker nodes and 2 topics. This is used as a broker between producer and consumer to decouple the applications.

IBM Bluemix Analytics Engine which is based on Apache Hadoop and Spark is used to store the data in Hadoop Distributed file System and perform analytics.

AWS RDS is used to store the data after performing analytics in hive. The filtered crime records during the Spark Streaming are also stored in AWS RDS.

The NodeJS application is deployed on AWS EC2 instance which is available for users to register and subscribe for the crime alerts.

The following technologies are used in our project:

1. *Kafka*- Kafka is used as Real-time messaging queue between producer and consumer.
2. *Spark*- Spark streaming is used for reading data from Kafka and perform analytics.
3. *HDFS*- The data is stored in Hadoop distributed file System.

4. *Scala*- We have used Scala as the programming language for working on Apache Spark.
5. *Hive*- Hive is used as Data-Warehouse in our project.
6. *NodeJS*- Developed a web application using NodeJS for users to subscribe for alerts and admin to notify the alerts.
7. *MySQL*- Used MYSQL for storing and retrieving the data in AWS RDS.

#### IV. CONCLUSION

A crime alert system is an application that leverages data analytics on crime data and demographic data to provide solutions to the end users.

Crime analytics and alert system can also be used by the law enforcement officials to curb the crime rate by incorporating police force towards critical areas responsible for these crimes.

#### ACKNOWLEDGMENT

We would like to thank Prof. Rakesh Ranjan for his extensive support in all the phases of project implementation.

#### REFERENCES

- [1] California state demographic data:  
[https://data.ca.gov/search/field\\_topic/economy-and-demographics-3](https://data.ca.gov/search/field_topic/economy-and-demographics-3)
- [2] Crime data by city in United States:  
[https://catalog.data.gov/dataset?organization\\_type=City+Government#topic=cities\\_navigation](https://catalog.data.gov/dataset?organization_type=City+Government#topic=cities_navigation)
- [3] Crime Reporting data state-wise in United States:  
<https://www.ucrdatatool.gov/Search/Crime/State/StatebyState.cfm>
- [4] An introduction of elasticsearch integration with node JS:  
<https://www.compose.com/articles/getting-started-with-elasticsearchand-node>.