

# A Systematic Approach toward Description and Classification of Cybercrime Incidents

**D.Vembu**

Periyar Manniammai Institute of Science and Technology,  
Department of CSE  
[vembu65@gmail.com](mailto:vembu65@gmail.com)

**P.Akalya A.Dinesh Kumar  
Tharangini.G**

Periyar Manniammai Institute of Science and Technology,  
Department of CSE

**ABSTRACT**— This project deal with a crime detail analysis method. It is developed to makes seven modules. Crime denotes to include a computer and a network. The computer might have been used in the commission of a crime, or it may be the target. Normally we are collect criminal detail and crime details. Our database should be converted into a backup file. Export and import must be made with that file. We can find the output in two ways. Terminal and browser

**Index Terms** - Hadoop, data analysis, export and import cybercrime.

## 1. INTRODUCTION:

### 1.1 Hadoop

Our first think is big data, Right now we living in data world, so how we are getting that much of huge data, if we can see different types of data generation factors are there, that is sensors, cc cams, online shopping air lines, face book, hospitality data, social network, in this way we getting huge data, if we are living in 100% of data world and is 90% of data has been generate for the last two years and the remaining 10% of the data generate as the long back, why we are using, because number of resource, for example the year 1990, having one hard disks, hard disks capacity 1GB to 20GP, and your ram capacity is 64 to 128 mb and your reading capacity is 10 kbps, this is what 1990's then 2014 so what is our hard disks capacity is minimum 1tb and your ram capacity 4 to 16 gb and your reading capacity is minimum 100 mbps

### 1.2 Hive

Hive is a supported to the any type of data. for example structure data, semi structure data and unstructured data. hive is a processing structure data. hive is a keeping a data from RDMS data and HDFS (Hadoop Distributed File System)data. Hive is a non job programmer . Firstly we having mysql type query .its supported to the hiveql(Hive Query Language). It is a good project in apache software foundation it has been given by facebook

### 1.2 Mapper code

So what is Mapper code, it is a WordMapper java, it should be define as public class WordMapper Extend Map reduce base implement Mapper, if a class is not going to implement a mapper interface, then only it is called as a mapper class

## 2.OBJECTIVE

In this paper we are analyzing crimedata by using hadoop tool along with some hadoop cybercrime like hdfs, mapreduce, sqoop, hive and pig. By using these tools we can process no limitation of data, no data lost problem, we can get high throughput, maintence cost also very less and it is a opensoure software,it is compatible on all the platforms since it is Java based.

## 3.EXISTING SYSTEM

Existing concept deals with providing backend by using mysql which contains lot of drawbacks i.e data limitation is that processing time is high when the data is huge and once data is lost we cannot recover so thus we proposing concept by using Hadoop framework.

## 4.DRAWBACKS IN EXISTING SYSTEM

We can process limitation of data. We get results with take more time and maintence cost is very high.

## 5.PROPOSED SYSTEM:

Proposed concept deals with providing database by using hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster and we get results with less time, high throughput and maintence cost is very less and we are using joins, partitions and bucketing techniques in hadoop.

## 6. ADVANTAGES IN PROPOSED SYSTEM

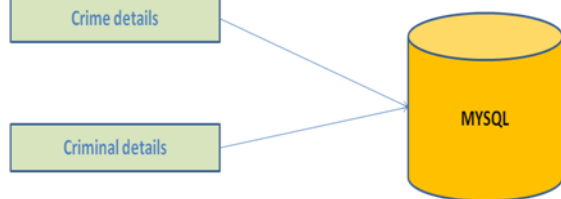
- No data loss problem
- Efficient data processing

## 7. METHODOLOGIES

- Data Preprocessing Module
- Data Ingestion Module With Sqoop
- Data Analytic Module With Hive
- Data Analytic Module With Pig
- Data Analytic Module With MapReduce
- Data Analytic Module With R
- Data Analytic Module With java generation

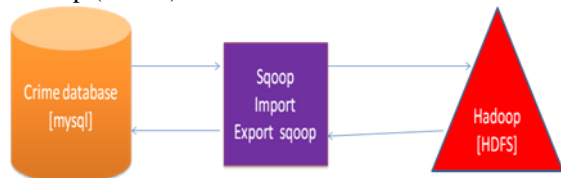
### Collect requirement dataset and Backup the file on mysql

We have to create a new table by right clicking on the database and selecting create new table, first column will be the Id which will be an integer, then all process will be completed, finally we got a backup file



### Data ingestion using sqoop

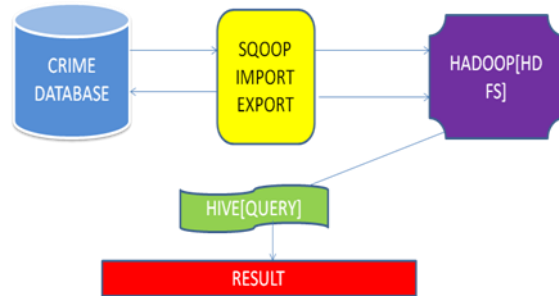
Data ingestion Module with Sqoop  
 Now we are ready with dataset. So now our aim is transfer the dataset into hadoop(HDFS), that will be happen in this module. Sqoop is a command-line interface application for transferring data between relational databases and Hadoop In this module we fetch the dataset into hadoop (HDFS) using sqoop Tool. Using sqoop we have to perform lot of the function, such that if we want to fetch the particular column or if we want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in hadoop (HDFS).



### Data processing Module with Hive

Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions.

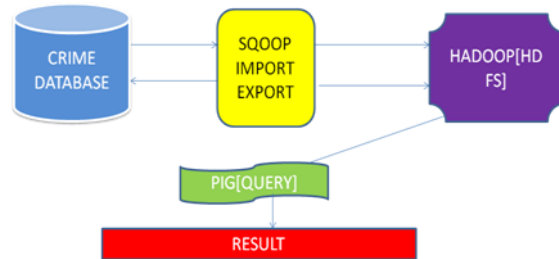
In this module we have to analysis the dataset using HIVE tool which will be stored in hadoop (HDFS). For analysis dataset HIVE using HQL Language. Using hive we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language.



### Data Analytic Module with Pig

Apache Pig is a high level data flow platform for execution Map Reduce programs of Hadoop. The language for Pig is pig Latin. Pig handles both structure and unstructured language. It is also top of the map reduce process running background.

In this module also used for analyzing the Data set through Pig using Latin Script data flow language. In this also we are doing all operators, functions and joins applying on the data see the result.



### Data Analytic Module with MapReduce

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. In this module also used for analyzing the data set using MAP REDUCE. Map Reduce Run by Java Program



## 8. ALGORITHM

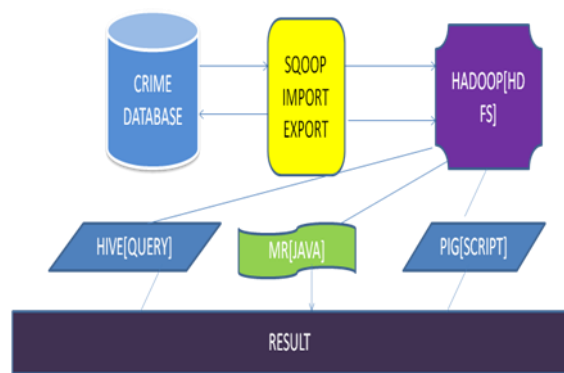
Generally Map Reduce paradigm is based on sending the computer to where the data resides!

Map Reduce program executes in three stages, namely map stage, shuffle stage, and reduce stage.

**Map stage :** The map or mapper job is to process the input data. Generally the input data is in the form of file or directory and is stored in the Hadoop file system (HDFS). The input file is passed to the mapper function line by line. The mapper processes the data and creates several small chunks of data.

**Reduce stage :** This stage is the combination of the Shuffle stage and the Reduce stage. The Reducer's job is to process the data that comes from the mapper. After processing, it produces a new set of output, which will be stored in the HDFS.

## 9.SYSTEM ARCHITECTURE & EXPLANATION



## 10. SOFTWARE REQUIREMENTS

### 10.1 H/W System Configuration

- Processor – Dual Core
- Hard Disk – 1 TB
- Memory – 4GB RAM
- Internet Connection

### 10.2 S/W System Configuration

- Windows/Linux
- VMware
- MYSQL

## 11. CONCLUSION

This paper illustrated the crime data analysis using Hadoop. The capability of big data will transform the way today's cyber crime to get criminal record. In the nearby future we will see implementation of big data analytics in crime data analysis. Big data provides security and privacy. This paper proposes a framework which is aiming that it will improve

the performance of MapReduce workloads and at the same time will maintain the fairness.

## ACKNOWLEDGMENT

It is our great pleasure to thank our assistant professor Ms.S.Karthigaiveni to encouraging us to do this paper presentation

## REFERENCE

- [1] D. L. Shinder and M. Cross, Scene of the Cybercrime. Burlington, MA, USA: Syngress, 2008.
- [2] FBI and NW3C. (May 22, 2015). 2014 Internet Crime Report. Accessed on May 17, 2016.
- [3] "IFCC 2002 Internet fraud report," FBI, Washington, DC, USA, and NW3C, Glen Allen, VA, USA, Tech. Rep. NCJ 194344, 2003.
- [4] PricewaterhouseCoopers. (Sep. 30, 2014). The Global State of Information Security Survey 2015—Managing Cyber Risks in an Interconnected World. Accessed on May 19, 2016.
- [5] Federal Criminal Police Office. (2009). Police Crime Statistics 2008. Accessed on May 18, 2016.
- [6] 2015 US State of Cybercrime Survey, PwC, London, U.K., Jul. 2015, accessed on May 18, 2016.
- [7] R. Anderson et al., "Measuring the cost of cybercrime," in The Economics of Information Security and Privacy. Berlin, Germany: Springer, 2013
- [8] L. D. Roberts, Cyber-Victimisation in Australia: Extent, Impact on Individuals and Responses. Hobart, TAS, Australia: Tasmanian Inst. Law Enforcement Stud., 2008.
- [9] K. M. Finklea and C. A. Theohary, Cybercrime: Conceptual Issues for Congress and U.S. Law Enforcement. Washington, DC, USA: Congressional Res. Service, 2012.
- [10] Council of Europe. (Nov. 23, 2001). Convention on Cybercrime. Accessed on Apr. 27, 2016.
- [11] Council of Europe. (Nov. 23, 2001). Summary to the Convention on Cybercrime (ETS N° 185) Treaty Office. Accessed on Dec. 18, 2015
- [12] Council of Europe. (Nov. 23, 2001). Explanatory Report to the Convention on Cybercrime (ETS N° 185) Treaty Office. Accessed on Dec. 18, 2015.
- [13] A. Murray, S. Zeadally, and A. Flowers, "An assessment of U.S. legislation on cybersecurity," in Proc. Int. Conf. Cyber Security Cyber Warfare Digit. Forensic (CyberSec), Kuala Lumpur, Malaysia, 2012
- [14] S. Gordon and R. Ford, "On the definition and classification of cybercrime," J. Comput. Virol., vol. 2, no. 1, pp. 13–20, 2006.
- [15] D. Wall, Cybercrime: The Transformation of Crime in the Information Age. Cambridge, U.K.: Polity, 2007.
- [16] U.S. House of Representatives. (1986). 18 U.S. Code §1030—Fraud and Related Activity in Connection With Computers, LII / Legal Information Institute. Accessed on May 27, 2016

**INTERNATIONAL JOURNAL FOR TRENDS IN ENGINEERING & TECHNOLOGY**  
**VOLUME 32 ISSUE 1 – AUGUST 2018 - ISSN: 2349 – 9303**

[17] CCIPS, “The National Information Infrastructure Protection Act of 1996—Legislative Analysis,” Comput. Crime Intellect. Property Section (CCIPS), United States Dept. Justice, Washington, DC, USA, Tech. Rep., Jun. 1998.

[18] I. Walden, Computer Crimes and Digital Investigations. Oxford, U.K.: Oxford Univ. Press, 2007.

[19] M. E. Kabay, “Understanding studies and surveys of computer crime,” in Computer Security Handbook. Hoboken, NJ, USA: Wiley, 2009.

[20] G. Kirwan and A. Power, The Psychology of Cyber Crime: Concepts and Principles. Hershey, PA, USA: IGI Glob., 2001.