

Credit Card Fraud Detection using Big Data Technologies

Lokesh¹, Rushikesh², Vaishnavi³, Prof. Sarika Aundhakar⁴

Students, Department of Computer Engineering^{1,2,3}

Professor, Department of Computer Engineering⁴

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *The use of credit cards is prevalent in modern day society. But it is obvious that the number of credit card fraud cases is constantly increasing in spite of the chip cards worldwide integration and existing protection systems. This is why the problem of fraud detection is very important now. Credit card fraud detection is the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect the fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on validating various Business Rules to Identify whether the transaction happened is Fraud/Genuine and report the same accordingly.*

Keywords: BIG DATA, HDFS, HIVE, H-BASE, KAFKA, UCL

I. INTRODUCTION

Cyber crime is any kind of crime that can be done in, with, or against networks and computer systems. Cyber crime is getting increased with the increasing threats due to online frauds and unethical hacking. With both information and cyber safety threat increasing, organizations must be ready to equip themselves with predicting and preventing cyber crime.

Cyber crime experts are using big data tools to identify the potential threats and detect cyber crime incidents like credit card frauds. Big data analytics is enabling companies to analyze voluminous amount of data they gather during financial transactions, any locale-specific data and others as well. Fighting cyber crime is of utmost importance today due to increased risk of cyber theft. Big data tools are being used to fight cyber attacks. Big Data analytics can help detect fraud and identify theft and can facilitate digital forensic analysis. In this paper, a brief survey is made about various techniques used in analysing big data to detect the frauds related to credit cards by analysing large set of data. One aim of this study is to identify the user model that best identifies fraud cases.

II. METHODOLOGY

Overall Processing Summary

Broadly classifies into Batch & Streaming Processing

2.1 Batch Processing

- Historical Card Transactions Dataset has to be loaded from Excel sheet to local system, from there to HDFS. Further this data will be moved to RDBMS using sqoop export Utility and to hive Hbase table.
- Member Score & member details data have been loaded to hive tables.
- Card Lookup Table is generated with card Transactions and Member Data.

2.2 Streaming Processing

- Data arrives via Kafka Topics and post validating in kafka based on Business rules marked as Fraud/Genuine and posted to card_transactions and card_lookup tables.

2.3 System Implementation Plan

Let's divide implementations process of project into sections and then in tasks.

Task 1- Copy "card_transactions.csv" file from local system to HDFS.

Table Creation

Task 2- Create the "card_transactions" table in MySQL based on the card_transactions.csv file Structure.

Task 3- Do a sqoop export to the database for card_transactions.csv and delete the file from HDFS

Task 4- On "member_score" and "member_details" create a normal hive external table. Task 5- Create a special "card_transactions" Hbase table managed by Hive.

Task 6- Create a Hbase "lookup" table with columns - member_id, card_id, UCL, timestamp, zipcode, credit_score.

Batch Processing

Task 7- Sqoop import member_score from AWS-RDS to Hive. (Full load import, has to be refreshed every week)

Task 8- Sqoop import member_details from AWS-RDS to Hive.

Task 9- Sqoop import card_transactions to HDFS from MySQL. (This is a one-time full load activity)

Scheduling

Task 10- Schedule a sqoop import job using Airflow to import member_score from AWS-RDS to Hive on a full- Load

Task 11- Schedule a sqoop import job using Airflow to import member_details from AWS- RDS to Hive on an incremental append mode for every 8hrs.

Integration

Task 12- Spark-HBase Integration

- For populating the card_transactions table.
- For populating the look_up table.

Task 13- Spark-Hive Integration for spark stream processing.

Task 14- Access the hive tables using apache spark and calculate the UCL.

Streaming:

Task 15- Producer to create the transactions in JSON format, to be added and queued in Kafka topics.

Task 16- Spark structured streaming program as a Consumer that will consume the data from the kafka topics.

Task 17- Retrieve the timestamp and zipcode of the last transaction of each card. Task 18- Processing in Spark Streaming – (Validating Rules)

- Validating RULE 1 -> "credit_score > 200"
- Validating RULE 2 -> "transaction amount <= UCL"
- Validating RULE 3 -> "zipcode distance within threshold"

Task 19- Based on the above rules, the entire transaction along with status should be updated .

Task 20- Schedule a job for validating rules by comparing the incoming data from the POS terminals in JSON format with the values in the lookup table.

Task 21- If the transaction was marked genuine, then we need to update the lookup table with the new timestamp and the zipcode.

Task 22- Schedule a job for populating the lookup table.

III. EXISTING SYSTEM

3.1 Credit Card Fraud Detection using Machine Learning

It is a process of data investigation by a Data Science team and the development of a model that will provide the best results in revealing and preventing fraudulent transactions. This is achieved through bringing together all meaningful features of card users' transactions, such as Date, User Zone, Product Category, Amount, Provider, Client's Behavioral Patterns, etc. The information is then run through a subtly trained model that finds patterns and rules so that

it can classify whether a transaction is fraudulent or is legitimate

IV. PROPOSED SYSTEM

4.1 Credit Card Fraud Detection using Big Data Technologies

The proposed credit card fraud detection system is being implemented with big data . It extends its feature and attempts to improve performance in addition with the three rules Which include:

1. Upper control limit.
2. Credit Score.
3. Zip Code Distance.

V. SYSTEM ARCHITECTURE

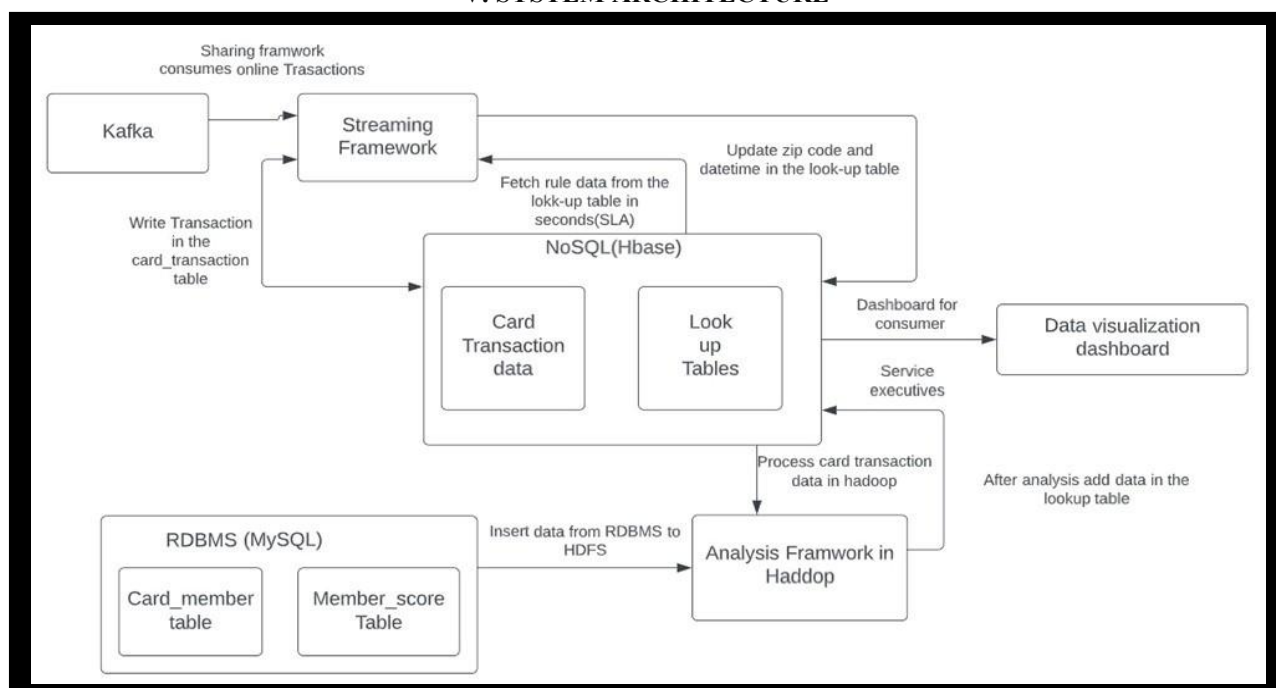


Figure: Fraud Detection Architecture

VI. CONCLUSION

An exceedingly proficient and exact credit card fraud detection framework is the need of great importance as a many transactions are being done each day. Thus a expansive measure of research is being done in this area furthermore, various systems are proposed conquered credit card extortion. The Fraud Detection System is additionally versatile for taking care of tremendous volumes of transactions. Our credit card fraud detection framework isn't taking long time what's more, having complex procedure to perform misrepresentation check like the existing framework and it gives preferable and quick come about over existing framework.

In this project, big data technologies like Spark, Kafka and zookeeper are being used for storing and processing of large amount of historical transactions of the users. Our result shows the effectiveness and correctness of the proposed system over a broad deviation of data.

REFERENCES

- [1]. Aditya B. Patel, Manashvi Birla, Ushma Nair , "Addressing Big Data Problem Using Hadoop and Map Reduce", International Conference on Engineering, 2012.
- [2]. Alessandro Guarino, "Digital Forensic as a Big Data Challenge", ISSE Securing Electronic Business Processes, 2013.

- [3]. Alexander, Rico Bergmann, Stephan, "The Stratosphere Platform for Big Data Analytics", the VLDB Journal, 2014
- [4]. R.Anbuvizhi, V.Balakumar, "Credit / Debit Card Transaction Survey Using Map Reduce in HDFS and Implementing Syferlock to Prevent Fraudulent", International Journal of Computer Science and network security, 2016.
- [5]. Anushree Naik, Kalyani Phulmamdikar, Shreya Pradhan, Sayali Thorat, Prof. Sachin V. Dhande, "Real time Credit card transaction analysis", International Engineering Research Journal (IERJ), Vol 1, Issue 11, 2016.
- [6]. Brian Ye, Anders Ye, "Exploring the Efficiency of Big Data Processing with Hadoop MapReduce", School of Computer Science and Communication (CSC), Royal Institute of Technology KTH, Stockholm, Sweden.
- [7]. Cameron S.D. Brown, "Investigating and Prosecuting Cyber Crime: Forensic Dependencies and Barriers to Justice", International journal of Cyber Criminology, 2015.
- [8]. D. J. Weston, D. J. Hand, N. M. Adams, C. Whitrow, P. Juszczak, "Plastic card fraud detection using peer group analysis", Advances in Data Analysis and Classification, 2008.
- [9]. Dr. Hugh J. Watson, "Big Data: Concepts, Technologies and Applications" [Online], Available: <http://www.watson-tutorial-big-data-businessanalytics-collaborative.pdf/>
- [10]. "Gartner Report: Big Data will Revolutionize Cyber Security in the Next Two Years" [online], Available: <http://cloudtimes.org/2014/02/12/gartnerreport-big-data-will-revolutionize-the-cybersecurityin-next-two-year/>
- [11]. "Hadoop, MapReduce, Spark, Flink" [Online], Available. <https://www.infoworld.com/article/2919602/flink-hadoops-new-contender-for-mapreduce-spark.html>
- [12]. J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters", Communications of the ACM, vol 51, pp. 107-113, 2008.
- [13]. Katarina Grolinger, Michael Hayes, Wilson A. Higashino, Alexandra L'Heureux, David S. Allison, and Miriam A.M. Capretz, "Challenges for MapReduce in Big Data", Proc. of the 10th 2014 world congress on services, 2014.
- [14]. K. R. Seeja and Masoumeh Zareapoor, "Fraud Miner: A Novel Credit card fraud detection model based on Frequent Item set Mining", The Scientific World Journal, 2014.
- [15]. Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica, "Spark: cluster computing with working sets", HotCloud'10 Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, 2010
- [16]. M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics", Journal of Big Data, 2015.
- [17]. Palak Gupta, Nidhi Tyagi, "An Approach towards Big Data –A Review", International Conference on Computing, Communication and Automation (IEEE), 2015.
- [18]. Paris Carbone, Gyula, Stephen, Seif, Kostas, "Light weight Asynchronous Snapshots for Distributed data flows", Cornell University Library, 2015.
- [19]. Palak Gupta, Nidhi Tyagi, "An Approach towards Big Data –A Review", International Conference on Computing, Communication and Automation (IEEE), 2015.
- [20]. J R.J.Bolton, D.J. Hand, "Unsupervised profiling methods for fraud detection", Proceedings of the VII Conference on Credit Scoring and Credit Control, 2001.