



Bigdata Analytics In Demonization

M. Kavitha*

Department of Information Technology, SRM University, Chennai, Tamil Nadu, India.

*Corresponding author's E-mail: kavitha.mu@ktr.srmuniv.ac.in

Received: 07-01-2017; Revised: 28-02-2017; Accepted: 15-03-2017.

ABSTRACT

The demonetization of ₹500 and ₹1000 notes was a policy endorsed by the Government of India on 8 November 2016. These actions were taken to aim at controlling black money, ensuring transparency, and annihilating graft from the economy. The result of this announcement, lead to lot of e-transaction by people. This e-transaction can lead to security breach. This paper suggests some ways to handle data by different banks, without breaking the security. Big Data Analytics can exhume black money being scattered in the Indian economy. Bigdata analytics can be used to uncover unknown correlations and hidden patterns, which helps the government in controlling the floating black money.

Keywords: demonetization, e-transaction, security, bigdata analytics.

INTRODUCTION

The bill was made by the Prime Minister Mr. Narendra Modi, that use of all ₹500 and ₹1000 banknotes would be invalid from midnight.

Big Data tools can run over complex and enormous data sets, which is not possible with traditional data-processing methods.

As per Mr. Modi, bills have already passed to all banks stating that accounts have to be linked with the Aadhar cards or the PAN card itself makes the transactions visible. But still without tools data Analysis is quite difficult.

Millions of millions of transactions undertaken every day will therefore throw-up insightful data that will form the embankment of information for the government to track. This is where big data and analytics will come into play¹.

The result of analysis not only helps to identify just individuals but also companies or financial institutions and other individuals involved in tax dodging³, money-laundering, identify holdups or other illegal activities.

Security Breach

Study says, the recent ATM hacking in India, which compromised over 30 lakh debit cards, resulted in a loss of confidence in digital payments by Indians everywhere.

Challenges in Analytics

Predictive analytics is projection of what might happen in future, because they are probabilistic in nature.

While dealing with banking transaction data, we have certain Management challenges to be addressed like data privacy, security, governance and ethical issues.

Data Privacy

How data privacy is preserved? What Level of privacy?

Data Security

How secure is my enterprise data? How my data is used?

Data Ethics

What legal issues is in my data? What about my ethical concern?

Data Governance

How quality is my result? Any risk in applying any model?

How Mapreduce helps in Analysis

Big Data Analytics tools provide hints to economic abuses in real time. The tools can be used to detect money corruption in India. By default every notes have denomination, country identifier, unique serial number and mechanism for fake prevention.

All banks and financial institutions use cash counting machines, which can also be slightly equipped with enhanced tools to detect and store serial numbers of currency nodes, regional information and details of mechanism for fake prevention.

Data analytics can then process the data sets to uncover **hidden patterns and unknown correlations**.

Algorithms can be used to find interpret where currencies comes from? And make predictions on patterns of money where cash hoarding take place.

With the help of parameters like **geographical location, serial numbers, individual account**, amount of transaction etc., algorithm can find patterns.

<location>Chennai</location>

<serial number>152587</serial number>

<account>0 96858631412582</account>

<amount-of-transaction>100000</amount-of-transaction>



Table 1: XML format of transaction the MapReduce algorithm performs the following actions –

• Tokenize –	Tokenizes the transaction into maps of tokens and writes them as key-value pairs.
• Filter –	Filters unwanted information from the maps of tokens and writes the filtered maps as key-value pairs.
• Count –	Generates a token counter per word.
• Aggregate Counters –	Prepares an aggregate of similar counter values into small manageable units.

Dataset

The data once provided by banks can be used for analysis to find out unusual movement in currency flow at any time. Sample analysis can be done with the help of free data set available.

Steps Involved in Processing**Identification and gathering of data**

Identify which type of banking data set is used for research ex. Transaction of a person on a monthly basis.

Store the big data

Where and how secure is the transaction details are stored without violating the privacy of the customer.

Perform targeted analytics

Targeted analysis is performed on dataset.

Metrics Used**Term Frequency (TF)**

It measures how frequently a particular term occurs in a document. It is calculated by the number of times a word appears in a document divided by the total number of words in that document.

$$TF = \frac{\text{number of times the transaction more than 10,0000}}{\text{Total number of transactions}}$$

CONCLUSION

People have been forced to go finally cashless, everyone even poor and those in rural areas who have little or no understanding of the online world, or the risks involved with digital payments.

In a digitised economy, every transaction will leave its own trace.

With the help of trace we can perform our research either on data analysis or in security.

REFERENCES

1. IDC iView "Extracting Value from Chaos," June 2011, sponsored by EMC.
2. IDC iView "Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," December 2012, sponsored by EMC.
3. IBM Global Business Services an Executive Report "Analytics: The real-world use of big data".
4. Rich Miller. "The Lessons of Moneyball for Big Data Analysis". www.datecenterknowledge.com. Retrieved 12 December 2015.

Source of Support: Nil, Conflict of Interest: None.

