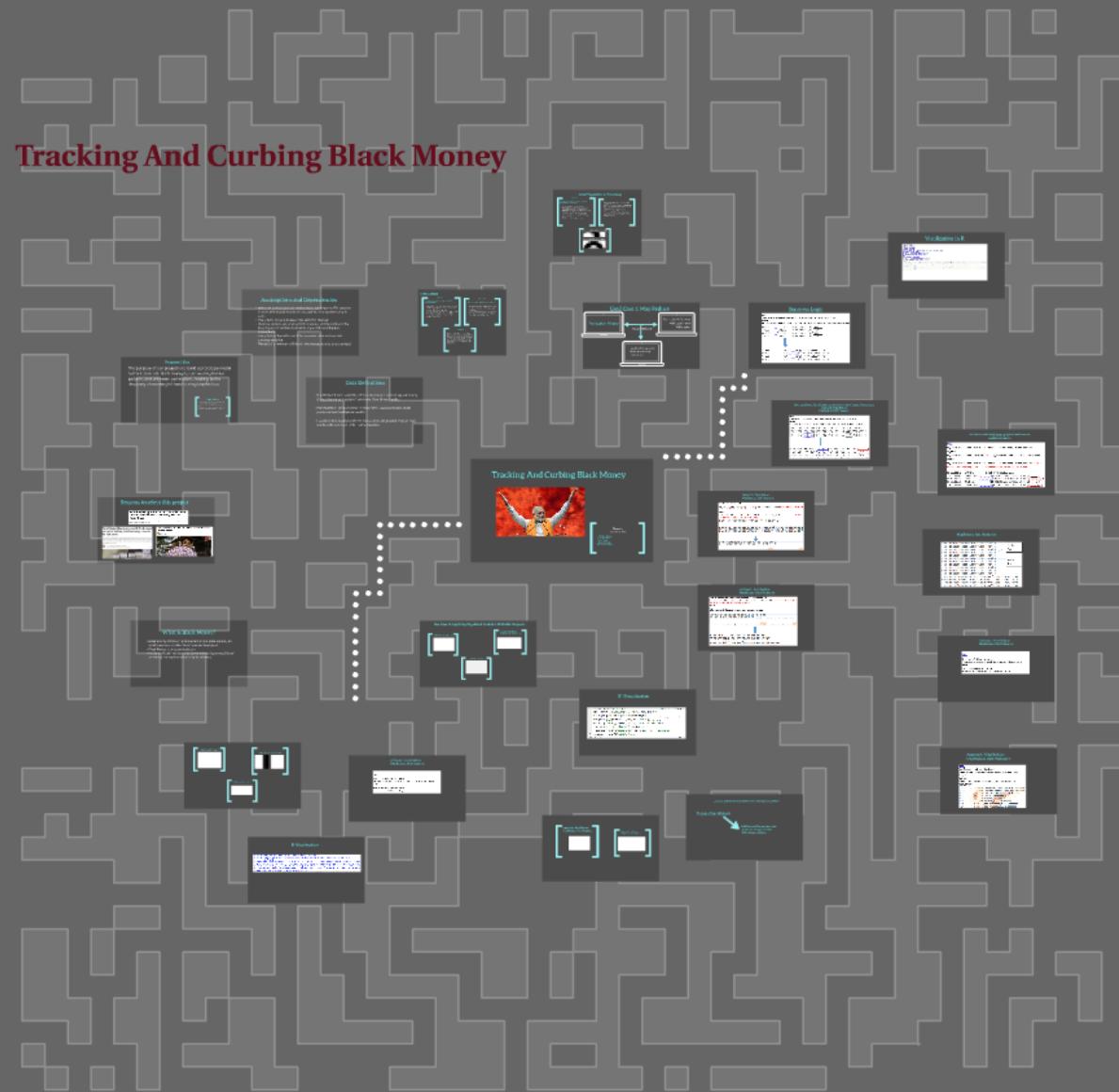


Tracking And Curbing Black Money



Tracking And Curbing Black Money

Purpose Idea:
The purpose of our project is to build a prototype model to track and curb black money by uncovering hidden patterns and unknown correlations; leading to the discovery of meaningful trends using MapReduce.

Assumptions and Dependencies

- All public transactions are considered as banking and ATM systems.
- All types of black money or electronic money do not exist in India.
- India has a unique currency code.
- There are no multiple bank accounts under one Aadhar card.
- Every Aadhar will receive the associated transaction and vice versa.
- Every Aadhar will have a unique ID.
- Distribution of Aadhar ID for the transaction to be identified.

UML Class Diagrams



Tracking And Curbing Black Money



Prime Minister
Narendra Modi

Reasons to select this project



What is Black Money?

- Black money refers to funds earned through black money, an illegal business, and other illegal activities.
- Black Money is also known as dirty money.
- Black money is often used to finance political parties.

R Visualization



Approach - Map Reduce



R Visualization

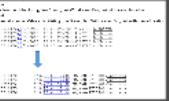


R Visualization

Used Case 1: Map Reduce



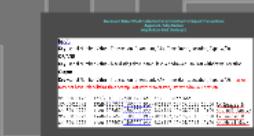
Business Logic:



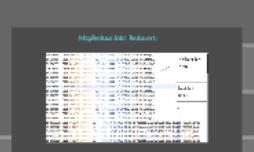
Visualization in R



Redundant Data in Black Money Detection



Redundant Data in Black Money Detection



MapReduce Job Redundancy



Approach - Map Reduce



Approach - Map Reduce

Transaction History



Transaction History



Transaction History

Tracking And Curbing Black Money



Team 15
Team Members

Akshay Nakhawa
Pranay Sampat
Ruchi Goyal
Maansi Chandira
Priyanka Bandekar

Team 15

Team Members

Akshay Nakhawa

Pranay Sampat

Ruchi Goyal

Maansi Chandira

Priyanka Bandekar

What is Black Money?

- Black money refers to funds earned on the black market, on which income and other taxes have not been paid.
- Black Money is unaccounted cash.
- Holders of black money try to convert it into legitimate ('clean' or 'white') money through money laundering.

Reasons to select this project

As India struggles for its next Rs 2,000, new currency seized across country can fill gunny bags

The new currency notes was seized from Haryana, Gujarat and Maharashtra and several people were arrested in connection with the seizure.

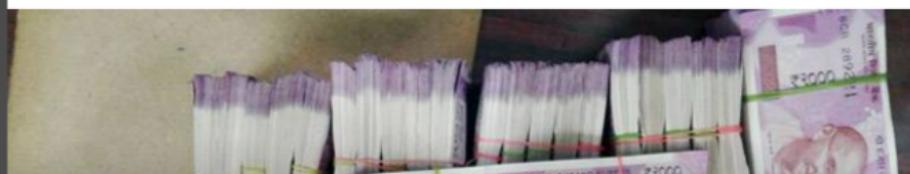
Tamil Nadu: New notes worth Rs 24 crore seized in Vellore, total recovery rises to Rs 166 crore

The Income Tax department today made fresh seizure of Rs 24 crore in new notes in Chennai. Officials said the fresh seizure of new currency, in Rs 2000 notes, was made by the sleuths from a car in Vellore on the insistence of the accused presently being interrogated in the case.



PTI | Posted by Yashaswani Sehrawat
Chennai, December 10, 2016 | UPDATED 14:18 IST

A + A -



Cash crunch? Rs 242 crore in new currency seized after demonetisation

INDIA Updated: Dec 10, 2016 01:14 IST
 Appu Esthose Suresh
Hindustan Times



Proposed Idea

The purpose of our project is to build a prototype model to track and curb black money by uncovering hidden patterns and unknown correlations, leading to the discovery of meaningful trends using MapReduce.

Project Objective

- To predict from where the black money is coming from and possible locations where cash hoarding may take place
- Detect the prospective black money transaction and suspected Pan Card Nos.
- The results from our project will in turn will be helpful for Tax department and other government agencies to catch black money holders.

Project Objective

- To predict from where the black money is coming from and possible locations where cash hoarding may take place
- Detect the prospective black money transaction and suspected Pan Card Nos.
- The results from our project will in turn will be helpful for Tax department and other government agencies to catch black money holders.

Data Definitions

- Blacklisted Notes: Serial No of the notes which did not appear in any transaction in our system from more than three months
- PAN Number: Unique identifier [Like SSN] associated with single person across multiple accounts.
- Location: Geo locations where transactions originated, may or may not be different than PAN – home location.

Assumptions and Dependencies

- Only cash transactions are considered at banking and ATM systems as most of the black money or unaccounted money exists only in cash.
- Every bank account is associated with Pan Number.
- Banking systems are enabled with Scanning and Recording of the Currency serial number involved in any withdrawal/deposit transactions.
- Every transaction will record the associated Pan number and currency serial No
- Threshold is minimum of 5 lac for the transactions to be scrutinized.

USE CASES

Use Case 1:

Detecting Transactions with more than 70% BlackListed Notes

- Process all the transaction and the associated currency no's and detecting if the deposited transaction has more than 70% of blacklisted notes .
- Deriving a relationship between the Depositor and last known Withdrawers of those BlackListed Notes involved in such transaction along with their corresponding locations.

Use Case 2:

Detecting Probable Black Money Transaction

- Detect deposit transactions in which 70 % of currency serials repeat from a withdrawal transaction.
- In other words it attempts to detect probable cash transactions between individuals by detecting common serial numbers.

Use Case 3

This use case attempts to apply Page Rank algorithm to our project. We apply Page Rank to the output from the Use case 1 and Use case 2, and try to rank the suspects depending upon number of black money transactions they are involved and the rank of the individuals, it is involved with.

Use Case 1:

Detecting Transactions with more than 70% BlackListed Notes

- Process all the transaction and the associated currency no's and detecting if the deposited transaction has more than 70% of blacklisted notes .
- Deriving a relationship between the Depositor and last known Withdrawers of those BlackListed Notes involved in such transaction along with their corresponding locations.

Use Case 2:

Detecting Probable Black Money Transaction

- Detect deposit transactions in which 70 % of currency serials repeat from a withdrawal transaction.
- In other words it attempts to detect probable cash transactions between individuals by detecting common serial numbers.

Use Case 3

This use case attempts to apply Page Rank algorithm to our project. We apply Page Rank to the output from the Use case 1 and Use case 2, and try to rank the suspects depending upon number of black money transactions they are involved and the rank of the individuals, it is involved with.

Data Preparation & Processing

Use Case 1:

Detecting Transactions with more than 70% BlackListed currency Serials

- Financial Records are not available publically.
- Currency Data with the cash hoarding not available.
- To design data which can incorporate 70% of serial number in one deposit transaction which were involved in another withdrawal transaction.
- Designing data such that the time-stamp of deposit should be later than withdrawal [which will always be the case in real-time.
- Generating transactions above the threshold.

- Generated black listed currency data, from official currency table.
- Designing black listed currency with PAN number and location of the person/party involved in holding black listed currency.
- Generating transactions above threshold
- Designing transaction data which will involve 70 % of same serial number currency in deposit transaction that was previously involved in withdrawal transaction by some another PAN.

Input File -01 : Black Listed Currency – 0.1 Million

Currency Serial Number	Timestamp	Last Withdrawal_PAN	Location
SN-50011172	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011173	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011174	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011175	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011176	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011177	1/3/2016 12:00 PM	PAN-9773780	Ludhiana

Input File 02 : Transaction File - 1 Million

TRANSACTION_ID	AMOUNT	PAN NUMBER	TIMESTAMP	LOCATION	TYPE	SERIAL_NUMBER
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111772
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111773
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111774
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111775
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111776
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111777
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111778
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111779
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111780
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-500111781

Use Case 1:

Detecting Transactions with more than 70% BlackListed currency Serials

- Financial Records are not available publically.
- Currency Data with the cash hoarding not available.
- To design data which can incorporate 70% of serial number in one deposit transaction which were involved in another withdrawal transaction.
- Designing data such that the time-stamp of deposit should be later than withdrawal [which will always be the case in real-time.
- Generating transactions above the threshold.

- Generated black listed currency data, from official currency table.
- Designing black listed currency with PAN number and location of the person/party involved in holding black listed currency.
- Generating transactions above threshold
- Designing transaction data which will involve 70 % of same serial number currency in deposit transaction that was previously involved in withdrawal transaction by some another PAN.

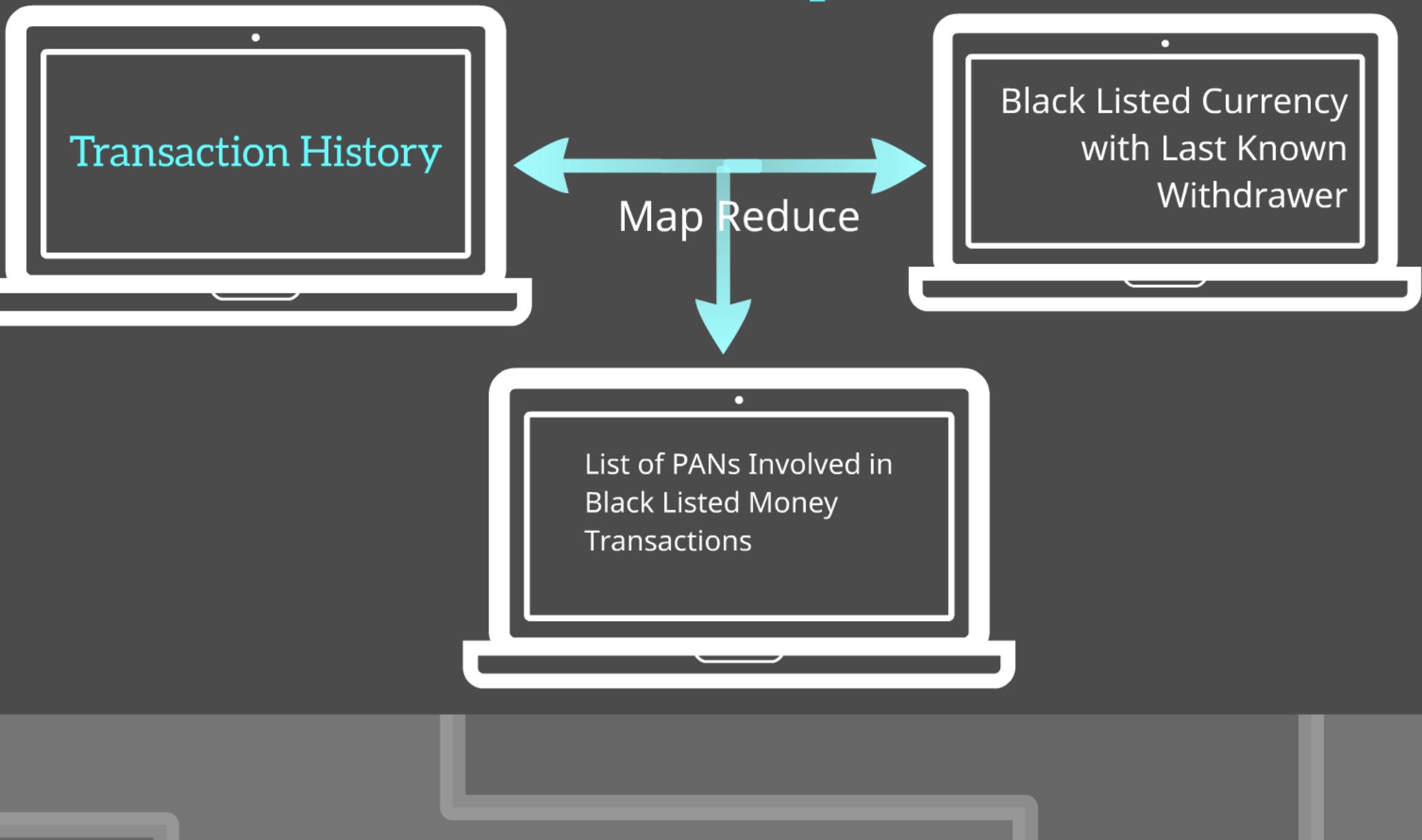
Input File -01 : Black Listed Currency – 0.1 Million

Currency Serial Number	Timestamp	Last Withdrawal_PAN	Location
SN-50011172	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011173	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011174	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011175	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011176	1/3/2016 12:00 PM	PAN-9773780	Ludhiana
SN-50011177	1/3/2016 12:00 PM	PAN-9773780	Ludhiana

Input File 02 : Transaction File - 1 Million

TRANSACTION_ID	AMOUNT	PANNUMBER	TIMESTAMP	LOCATION	TYPE	SERIAL_NUMBER
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011772
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011773
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011774
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011775
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011776
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011777
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011778
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011779
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011780
TRAN-1263526	500000	PAN-9826117	12/05/2016 10:00 AM	Mumbai	D	SN-50011781

Used Case 1: Map Reduce



Business Logic

Input -

Serial Number, blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location

Output

Key – serial Number **Value** - blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location

SN-50011172, 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011173, 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011174, 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011175, 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011176, 1/3/2016 12:00 PM, PAN-9773780, Ludhiana



SN-50011172 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011173 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011174 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011175 1/3/2016 12:00 PM, PAN-9773780, Ludhiana
SN-50011176 1/3/2016 12:00 PM, PAN-9773780, Ludhiana

Use Case1: Detect Black Listed Currency involved in Deposit Transactions

Approach : Map Reduce

MapReduce Job1 Mapper2:

Input –

Transaction ID, Amount, PAN, Time Stamp, Location, Type, currency Serial Number

Output

Key –serial Number **Value** - Transaction ID, Amount, PAN, Time Stamp, Location, Type(W/D)

```
TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D,SN-50011772
TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D,SN-50011773
TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D,SN-50011774
TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D,SN-50011775
TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D,SN-50011776
```



```
SN-50011772    TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D
SN-50011773    TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D
SN-50011774    TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D
SN-50011774    TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D
SN-50011774    TRAN-1263526,500000,PAN-9826117,12/05/2016 10:00 AM,Mumbai,D
```

Use Case1: Detect Black Listed Currency involved in Deposit Transactions

Approach : Map Reduce

MapReduce Job1 Reducer1:

Input :

Key –serial Number **Value** - Transaction ID, Amount, PAN, Time Stamp, Location, Type(W/D)

OR/AND

Key –serial Number **Value** - blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location

Output

Key –serial Number **Value** - Transaction ID, Amount, PAN, Time Stamp, Location, Type(W/D),**(may or may not have) blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location**

SN-50011172 1/3/2016 12:00 PM, PAN-9773780, Ludhiana

SN-50011772 TRAN-1263526, 500000, PAN-9826117, 12/05/2016 10:00 AM, Mumbai, D

SN-50011772 TRAN-1263524, 600000, PAN-9826119, 12/05/2015 10:00 AM, Ludhiana, W

SN-50011772 TRAN-1264526, 500000, PAN-9826127, 02/05/2016 10:00 AM, Panjim, D

SN-50011772 TRAN-1266626, 505000, PAN-2226117, 12/06/2016 10:00 AM, Bhatinda, W

MapReduce Job1 Reducer1:

SN-306999	TRAN-36465701,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
SN-50011253	TRAN-5432109,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011254	TRAN-5432109,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011255	TRAN-5432109,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011256	TRAN-5432109,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011257	TRAN-5432109,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi

Nonblack listed
Notes

Black listed
Notes

Approach : Map Reduce MapReduce Job2 Mapper1:

Input (Output of previous stage)

Key—serial Number **Value** - Transaction ID, Amount, PAN, Time Stamp, Location, Type(W/D),(may or may not have) blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location

Output

Key: Transaction Id **Value:** serial Number, Amount, PAN, Time Stamp, Location, Type(W/D),(may or may not have) blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location

```
SN-306999 TRAN-36465701,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
SN-306999 TRAN-36465701,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
SN-306999 TRAN-36465701,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
SN-50011253 TRAN-5432109,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011254 TRAN-5432109,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011255 TRAN-5432109,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011256 TRAN-5432109,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011257 TRAN-5432109,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
SN-50011258 TRAN-5432109,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
```



```
TRAN-5432109| SN-306999,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
TRAN-5432109 SN-306999,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
TRAN-5432109 SN-306999,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D
TRAN-5432109 SN-50011253,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
TRAN-5432109 SN-50011253,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
TRAN-5432109 SN-50011253,10000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
```

Approach : Map Reduce MapReduce Job2 Reducer1:

Input: Transactions and corresponding all the serial numbers (black listed or not)

Key: Transaction Id **Value:** serial Number, Amount, PAN, Time Stamp, Location, Type(W/D), (may or may not have) blacklisting time Stamp, last Withdrawer Pan, last withdrawal Location

Output:

Key: Depositor PAN, Depositor Location

Value: transaction Id, amount, last Withdrawer PAN, last withdrawer Location

```
|  
TRAN-5432109| SN-306999,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D  
TRAN-5432109 SN-306999,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D  
TRAN-5432109 SN-306999,10000000,PAN-0637446666,12/12/2016 11:03 AM,Panaji,D  
TRAN-5432109 SN-50011253,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi  
TRAN-5432109 SN-50011253,1000000,PAN-3258091,29/06/2016 09:35 AM,Kolkata,D,29/05/2016 3:00 PM,PAN-7011153,Kochi
```



```
PAN-3258091,Mumbai TRAN-4524241,1366000,PAN-01926673,Mumbai  
PAN-3258091,Kolkata TRAN-5432109,1000000,PAN-01926673,Mumbai  
PAN-3258091,Kolkata TRAN-5432109,1000000,PAN-7011153,Kochi  
PAN-3258091,Kolkata TRAN-5432109,1000000,PAN-10009831,Surat  
PAN-9826117,Delhi TRAN-7564528,1476000,PAN-145555551,Visakhapatnam  
PAN-00893897,Asansol TRAN-7960870,1500000,PAN-657546432,Amritsar
```

Approach : Map Reduce MapReduce Job3 Mapper1:

Input :

Key: Depositor PAN, Depositor **Location**

Value: transaction Id, amount, last Withdrawer PAN, last withdrawer Location

Output :

Key: Depositor PAN, Depositor **Location**

Value: amount, last Withdrawer PAN, last withdrawer Location

Approach : Map Reduce MapReduce Job3 Reducer1:

Input :

Key: Depositor PAN, Depositor Location

Value: amount, last Withdrawer PAN, last withdrawer Location

Output:

Key-none

Value- PAN, LOCATION, withdrawer PAN, LOCATION, AMOUNT |

FINAL OUTPUT

```
PAN-00893897,Asansol,PAN-804877661,Srinagar,1500000
PAN-00893897,Asansol,PAN-657546432,Amritsar,1500000
PAN-283632,Pune,PAN-10611411,Bhubaneswar,600000
PAN-283632,Pune,PAN-394743,Baharanpur,600000
PAN-283632,Pune,PAN-849298,Kolkata,600000
PAN-283632,Pune,PAN-516965,Asansol,600000
PAN-3258091,Kolkata,PAN-10009831,Surat,1000000
PAN-3258091,Kolkata,PAN-7011153,Kochi,1000000
```

Visulization in R

```
> library(maps)
> library(dplyr)
> library(geosphere)
> library(magrittr)
> library(leaflet)
> transaction<- read.csv("c:/users/priya/Desktop/trans.csv")
> dfPoints <- data.frame(
+   long = transaction$Transaction.Long,
+   lat = transaction$Transaction.Lat
+ )
> dfPoints1 <- data.frame(
+   long = transaction$L.Transaction.Long,
+   lat = transaction$L.Transaction.Lat
+ )
> #leaflet() %>% addTiles() %>%addMarkers(lng=transaction$Transaction.Long, lat=transaction$Transaction.Lat, popup=transaction$Total.Amount)%>%addMarkers(lng=transaction$L.Transaction.Long, lat=transaction$L.Transaction.Lat, popup=transaction$Total.Amount)%>%addPolylines(data = dfPoints, lng = ~long, lat = ~lat)%>%addPolylines(data = dfPoints1, lng = ~long, lat = ~lat)
> #using addrectangle
> leaflet() %>% addTiles() %>%addMarkers(lng=transaction$Transaction.Long, lat=transaction$Transaction.Lat, popup=transaction$Transaction.Location,)%>%addMarkers(lng=transaction$L.Transaction.Long, lat=transaction$L.Transaction.Lat, popup=transaction$BlacklistedWithdrawal)
> |
```

Use Case2: Detecting Probable Black Money Transactions (Anomalies)

Transaction History



PAN Pair of a Transaction with
Amount having more than
70% of note similarity

Approach : Map Reduce MapReduce Job1 Mapper1:

```
Input - TransactionID, Amount, Pmt, Timestamp, Location, Type, currencySerialNumber  
Output  
Key - TransactionID Value - Account,Pmt,Timestamp,Location,Pmt,currencySerialNumber  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001312  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001314  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001315  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001316  
  
↓  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001320  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001321  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001322  
TRAN-0000001,100000,PAW-9773700,29/06/2016 09:10 AM,Mumbai,PAW,SH-7001323
```

Approach : Map Reduce MapReduce Job1 Reducer1:

```
Input : Key - TransactionID Value - TxnID,Amount,Pmt,TimeStamp,Location,Type,currencySerialNumber  
Output : Key - TransactionID Value - TxnID,Amount,Pmt,TimeStamp,Location,Type,currencySerialNumber,notesCount  
Overall Job Output:  
TRAN-0000001 TRAN-0000001,1000000,PAW-9773700,29/06/2016 09:00  
AM,Mumbai,PAW,SH-7001312,500  
TRAN-0000001 TRAN-0000001,1000000,PAW-9773700,29/06/2016 09:00  
AM,Mumbai,PAW,SH-7001314,500  
TRAN-0000001 TRAN-0000001,1000000,PAW-9773700,29/06/2016 09:00  
AM,Mumbai,PAW,SH-7001315,500  
TRAN-0000001 TRAN-0000001,1000000,PAW-9773700,29/06/2016 09:00  
AM,Mumbai,PAW,SH-7001316,500  
TRAN-0000001 TRAN-0000001,1000000,PAW-9773700,29/06/2016 09:00  
AM,Mumbai,PAW,SH-7001317,500  
TRAN-0000001 TRAN-0000001,1000000,PAW-9773700,29/06/2016 09:00  
AM,Mumbai,PAW,SH-7001318,500
```

Approach : Map Reduce

MapReduce Job1 Mapper1:

Input -

TransactionID, Amount, Pan, Timestamp, Location, Type, currencySerialNumber

Output

Key - TransactionID Value - Amount, Pan, Timestamp, Location, Type, currencySerialNumber

TRAN-0000001, 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001312

TRAN-0000001, 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001314

TRAN-0000001, 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001315

TRAN-0000001, 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001316



TRAN-0000001 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001316

TRAN-0000001 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001317

TRAN-0000001 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001318

TRAN-0000001 1000000, PAN-9773700, 29/06/2016 09:00 AM, Mumbai, W, SN-7001319

Approach : Map Reduce

MapReduce Job1 Reducer1:

Input :

Key – TransactionID Value - TxnID,Amount,PAN,TimeStamp,Location,Type,currencySerialNumber

Output

Key – TransactionID Value - TxnID,Amount,PAN,TimeStamp,Location,Type,currencySerialNumber,notesCount

Overall Job1 Output:

TRAN-0000001	TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00
AM,Mumbai,W,SN-7001312,500	
TRAN-0000001	TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00
AM,Mumbai,W,SN-7001314,500	
TRAN-0000001	TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00
AM,Mumbai,W,SN-7001315,500	
TRAN-0000001	TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00
AM,Mumbai,W,SN-7001316,500	
TRAN-0000001	TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00
AM,Mumbai,W,SN-7001317,500	
TRAN-0000001	TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00
AM,Mumbai,W,SN-7001318,500	

MapReduce Job2 Mapper1:

Input (Output of previous stage)	
Key - TransactionID	Value - Tndt,Amount,STN,TimeStamp,Location,Type,currency,Geo,Wanderer,Notes,Cust
Output	
Key: Sales Number	Value: Tndt,Amount,STN,TimeStamp,Location,Type,currency,Geo,Wanderer,Notes,Cust
TRSD-00000001	STAN-00000001-PAT-5773700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800
TRSD-00000002	STAN-00000002-PAT-5773700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800
TRSD-00000003	STAN-00000003-PAT-5773700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800
	
TRSD-00000004	STAN-00000004-LAUSDNO,PAW-#73700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800
TRSD-00000005	STAN-00000005-LAUSDNO,PAW-#73700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800
TRSD-00000006	STAN-00000006-LAUSDNO,PAW-#73700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800
TRSD-00000007	STAN-00000007-LAUSDNO,PAW-#73700,23/06/2016 09:00 AM,Hanover,W,STN-7001318,800

MapReduce Job2 Reducer1:

```

Input: Key- Serial Number Value: TxnID,Amount,PAN,TimeStamp,Location,Type,notesCount
Output:
To make transaction pairs and output deposit and withdraw type locations and PAN number
Key: Transaction Pairs
Value: depAmount,depPAN,depTimeStamp,depLocation,notesCount,wPAN,wLocation
Final Output of Job2:

TRAN-1215687,TRAN-8343609      980000, PAN-0284283,01/01/2016 02:01
PN,Mumbai ,1900, PAN-1037227,Kolkata
TRAN-1215687,TRAN-5343609      980000, PAN-0284283,01/01/2016 02:01
PN,Mumbai ,1900, PAN-1037227,Kolkata
TRAN-1215687,TRAN-8343609      980000, PAN-0284283,01/01/2016 02:01
PN,Mumbai ,1900, PAN-1037227,Kolkata

```

MapReduce Job3 Mapper1 and Reducer1:

```
Input :
Key: Transaction PAN
Value: depAmount depPAN,depTimeStamp,depLocation,notesCount,wPAN,wLocation
Output:
Key: Depositor PAN, Location
Value: Withdrawer PAN, Location Amount
FINAL OUTPUT:

PAW-001612032_Ranjeet_PAN-30673705_Delhi_550000
PAW-001612032_Ranjeet_PAN-849369_Mohali_850000
PAW-001612032_Ranjeet_PAN-30673705_Guwahati_550000
```

MapReduce Job2 Mapper1:

Input (Output of previous stage)

Key - TransactionID Value - TxnID,Amount,PAN,TimeStamp,Location,Type,currency,SerialNumber,notesCount

Output

Key: Serial Number Value: TxnID,Amount,PAN,TimeStamp,Location,Type,notesCount

```
TRAN-0000001      TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00  
AM,Mumbai,W,SN-7001317,500  
TRAN-0000001      TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00  
AM,Mumbai,W,SN-7001318,500  
TRAN-0000001      TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00  
AM,Mumbai,W,SN-7001319,500
```



```
SN-7001312  TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00 AM,Mumbai,W,500  
SN-7001312  TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00 AM,Mumbai,W,500  
SN-7001315  TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00 AM,Mumbai,W,500  
SN-7001316  TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00 AM,Mumbai,W,500  
SN-7001317  TRAN-0000001,1000000,PAN-9773700,29/06/2016 09:00 AM,Mumbai,W,500
```

MapReduce Job2 Reducer1:

Input: Key- Serial Number Value: TxnID,Amount,PAN,TimeStamp,Location,Type,notesCount

Output:

To make transaction pairs and output deposit and withdraw type locations and PAN numbers.

Key: Transaction Pairs

Value: depAmount,depPAN,depTimeStamp,depLocation,notesCount,wPAN,wLocation

Final Output of Job2:

```
TRAN-1215657:TRAN-5343809      950000, PAN-0284283, 01/01/2016 02:01  
PM, Mumbai, 1900, PAN-1037217, Kolkata  
TRAN-1215657:TRAN-5343809      950000, PAN-0284283, 01/01/2016 02:01  
PM, Mumbai, 1900, PAN-1037217, Kolkata  
TRAN-1215657:TRAN-5343809      950000, PAN-0284283, 01/01/2016 02:01  
PM, Mumbai, 1900, PAN-1037217, Kolkata  
TRAN-1215657:TRAN-5343809      950000, PAN-0284283, 01/01/2016 02:01  
PM, Mumbai, 1900, PAN-1037217, Kolkata  
TRAN-1215657:TRAN-5343809      950000, PAN-0284283, 01/01/2016 02:01  
PM, Mumbai, 1900, PAN-1037217, Kolkata
```

R Visualization

```
> net2 <- graph_from_data_frame(d=links1, vertices=nodes1, directed=T)
> net4 <- network(links1, vertex.attr=nodes1, matrix.type="edgelist", loops=F, multiple=F, ignore.eval = F)
> par(mar=c(0,0,0,0))
> render.d3movie(net4, usearrows = T, displaylabels = F, bg="#111111", vertex.border="#ffffff", vertex.col = "green",
+ vertex.cex = (net4 %v% "size")/3, edge.lwd = (net4 %e% "weight")/3, edge.col =c("blue", "red")[(E(net1)$type=="black")+1], vertex.tooltip = paste( "<b>Pan No:</b>", (net4 %v% 'id') , "<br>" , "<b>Location:</b>" , (net4 %v% 'location')), edge.tooltip = paste("<b>Transaction:</b>", (net4 %e% 'type'), "<br>" , "<b>Amount:</b>" , (net4 %e% "Amount" )), launchBrowser=T, filename="Transaction-Network.html" )
```

Use Case 3: Applying PageRank to detect Probable Suspects

MapReduce Input to Iterator:

Approach : Map Reduce
MapReduce Output of Iterator:

MapReduce Output of Viewer:

MapReduce Input to Iterator:

```
PAN-000000014 : PAN-136736829
PAN-000000029 : PAN-10009831 PAN-30673705 PAN-849298 PAN-30673708 PAN-13581139 PAN-30673697 PAN-283632 PAN-311219
PAN-00000012  : PAN-000000029
PAN-0011728291 : PAN-29383850 PAN-29383849 PAN-29383848 PAN-29383847 PAN-29383846 PAN-29383845 PAN-29383844 PAN-29
PAN-001612032 : PAN-30673705 PAN-30673697 PAN-13581139 PAN-30673708 PAN-849298
PAN-001672661 : PAN-30673694 PAN-30673695 PAN-849298 PAN-30673705 PAN-786474849 PAN-99999999 PAN-88888888 PAN-777
PAN-00893897 : PAN-657546432 PAN-804877661
PAN-01926673 : PAN-768457543 PAN-13581139 PAN-01926673 PAN-10009831 PAN-657546432 PAN-804877661 PAN-768457543 PA
PAN-0284283 : PAN-0173712
PAN-03111990 : PAN-71381187
PAN-05031971 : PAN-31121991
PAN-0637446666 : PAN-145555551
PAN-08121990 : PAN-71381187
PAN-136736829 : PAN-80000001 PAN-99038397 PAN-83848937 PAN-8347468
PAN-14111990 : PAN-31121991
PAN-145555551 : PAN-657546432 PAN-768457543 PAN-804877661
PAN-18121958 : PAN-953990
PAN-19101992 : PAN-31121991
PAN-20102015 : PAN-71381187
PAN-24061987 : PAN-18121958
PAN-26101989 : PAN-71381187
PAN-283632 : PAN-10611411 PAN-953990 PAN-394743 PAN-516965 PAN-849298
PAN-29061990281 : PAN-88888888
PAN-29383854 : PAN-30673708 PAN-13581139 PAN-30673697 PAN-283632 PAN-31121991 PAN-0173712 PAN-31121991 PAN-88888
PAN-30673690 : PAN-283632 PAN-001612032
PAN-30673692 : PAN-18121958
PAN-30673693 : PAN-18121958
```

Approach : Map Reduce

MapReduce Output of Iterator:

PAN-30673702	0.9846461729300923	
PAN-30673703	2.4846461729300917:	PAN-71381187
PAN-30673704	2.4846461729300917:	PAN-71381187
PAN-30673705	1.7927763147365412	
PAN-30673706	2.2176386899173455:	PAN-786474849 PAN-18121958 PAN-283632
PAN-30673707	6.529092372670325	
PAN-30673708	4.426767218443426:	PAN-30673707 PAN-283632 PAN-786474849 PAN-99999999 PAN-88888888 PAN-7777777
PAN-31121991	7.0476684616626875:	PAN-001672661 PAN-001612032
PAN-3258091	1.4999999999999998:	PAN-10009831 PAN-7011153 PAN-01926673 PAN-01926673
PAN-394743	2.2973513073357963:	PAN-849298 PAN-71381187 PAN-516965 PAN-516965 PAN-13581139
PAN-516965	3.610952018345257:	PAN-30673695 PAN-10611411 PAN-18121958 PAN-849298 PAN-30673699
PAN-657546432	4.539001954441925:	PAN-768457543 PAN-145555551 PAN-0637446666
PAN-7011153	0.4293211269833985	
PAN-71381187	13.814240562792982:	PAN-30673707 PAN-18121958
PAN-748457543	4.405140356171724	
PAN-768457543	5.385376947245815:	PAN-748457543
PAN-77774647484	1.4999999999999998:	PAN-001672661
PAN-77777777	0.34854762506452336	
PAN-786474849	2.5487724544013455:	PAN-30673695 PAN-30673694 PAN-30673692 PAN-30673691
PAN-80000001	0.5896874999999999	
PAN-804877661	3.039001954441925	

MapReduce Output of Viewer:

-13.81424	PAN-71381187
-13.2682495	PAN-18121958
-7.1374335	PAN-1455555551
-7.0476685	PAN-31121991
-6.1926184	PAN-001672661
-5.496446	PAN-001612032
-5.385377	PAN-768457543
-4.7259855	PAN-30673697
-4.539002	PAN-657546432
-4.4267673	PAN-30673708
-4.1438594	PAN-283632
-3.610952	PAN-516965
-3.3598754	PAN-88888888
-3.0113277	PAN-000000029
-2.9834313	PAN-01926673
-2.8166823	PAN-0637446666
-2.775	PAN-0011728291
-2.775	PAN-136736829
-2.554086	PAN-30673699
-2.5487726	PAN-786474849
-2.484646	PAN-30673704
-2.484646	PAN-30673703
-2.2973514	PAN-394743
-2.2176387	PAN-30673706
-2.1158214	PAN-30673692
-2.0896876	PAN-83848937
-1.869091	PAN-30673700
-1.8485476	PAN-99999999
-1.8485476	PAN-26101989
-1.8485476	PAN-05031971
-1.6474218	PAN-29383854
-1.5	PAN-30673690
-1.5	PAN-998383983

R Visualization

```
1 nodes <- read.csv("/Users/Ruchi/Desktop/Final-Node.csv", header=T, as.is=T)
2 links <- read.csv("/Users/Ruchi/Desktop/Final-edges.csv", header=T, as.is=T)
3 net <- graph_from_data_frame(d=links, vertices=nodes, directed=T)
4 net3 <- network(links, vertex.attr=nodes, matrix.type="edgelist", loops=F, multiple=F, ignore.eval = F)
5 render.d3movie(net3, usearrows = T, displaylabels = F, bg="#111111",
6 vertex.border="#ffffff", vertex.col = "tomato", vertex.cex = (net3 %v% "size")/2,
7 edge.lwd = (net3 %e% "weight")/3, edge.col =c("blue", "red")[(E(net)$type=="black")+1]
8 ,vertex.tooltip = paste( "<b>Page Rank Weight:</b>",
9   (net3 %v% 'size') , "<br>","<b>Pan No:</b>", (net3 %v% 'id') ,
10  "<br>","<b>Location:</b>", (net3 %v% 'location')),
11  edge.tooltip = paste("<b>Transaction:</b>", (net3 %e% 'type'), "<br>","<b>Amount:</b>",
12  (net3 %e% "Amount" ) ),
13 launchBrowser=T, filename="BlackMon-Network.html" )
```

Tracking And Curbing Black Money

Purpose Idea:
The purpose of our project is to build a prototype model to track and curb black money by uncovering hidden patterns and unknown correlations; leading to the discovery of meaningful trends using MapReduce.

Assumptions and Dependencies

- All public transactions are considered as banking and ATM systems.
- All types of black money or electronic money do not exist in India.
- India has a unique currency code.
- There are no multiple bank accounts under one Aadhar card.
- Every Aadhar will receive the associated transaction and vice versa.
- Every Aadhar will have a unique ID.
- Distribution of Aadhar ID for the transaction to be identified.

UML Class Diagrams



Tracking And Curbing Black Money



Prime Minister
Narendra Modi

Reasons to select this project



What is Black Money?

- Black money refers to funds earned through black money, an illegal business, and other illegal activities.
- Black Money is also known as dirty money.
- Black money is often used to finance political parties.

R Visualization



Approach - Map Reduce (Machine Learning)



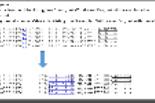
Data Preparation & Processing



Used Case 1: Map Reduce



Business Logic:



Visualization in R



Used Case 2: Map Reduce (Machine Learning)



MachineLearningJobResult:



R Visualization



Approach - Map Reduce (Machine Learning And Machine Learning)



Transaction History

