International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 1, January 2016

Credit And Debit Card Transaction Survey Using Map Reduce In Hdfs And Implementing Syferlock To Prevent Fraudulent

[1] R.Anbuvizhi [2] V.Balakumar
[1]Department of Computer Science and Engineering, [2] System Engineer
[1]Dhanalakshmi Srinivasan Engineering College [2]Tata Consultancy Service
[1]Perambalur, [2] Chennai,
TamilNadu, India
[1]r.anbuvizhi@gmail.com[1] [2]bala101289@gmail.com[2]

Abstract: **Big Data" is a term that has jumped overnight from its roots. It can be described as an innovative technique and technology to save, distribute, manage, visualize and analyze larger-sized data with extreme velocity and methods to manage unstructured and structure incapable amount of data. Nowadays fraud transactions in real time are increasing, so identifying large-scale patterns across many transactions or detecting anomalous behavior from an individual user can change the fraudulent in online transaction. Thus banks are turning to analytics to predict and prevent fraud in real times. In this project, proposed system deals with Credit and Debit card management to improve its scalability and efficiency in big data environment, it is implemented on (HDFS) Hadoop distributed file system. On the other hand, nowadays fraudulent and web attacker are increased to steel the password. In order to overcome those, proposed system implements SyferLock concept to add more security to Credit card transaction. The changes include a stronger user authentication tool, SyferLock that will require associates to follow a few new steps when logging into the network remotely. SyferLock patented, random number generation authentication solutions provide next generation One-Time Passwords/PINs (OTPs) for secure access to computers, networks and the Internet. SyferLock has engineered an enhanced authentication methodology and system using device-less OTPs that provides users with a simple, more secure way to access information leveraging their existing passwords. It uses a one-time password that expired once used. This added security strengthens the ability to protect the applications and systems.**

Index Items : **Analysis, Big Data, Credit Card, GridPin, Hadoop, Map Reduce, Report, Security, Syferlock**.

I. INTRODUCTION

A. Big Data

Big data is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large it is difficult to process using traditional database and software techniques. In most enterprise scenarios the volume of data is too big or it moves too fast or it exceeds current processing capacity. Despite these problems, big data has the potential to help companies improve operations and make faster, more intelligent decisions. Is Big Data a Volume or a Technology? While the term may seem to reference the volume of data, that isn't always the case. The term big data, especially when used by vendors, may refer to the technology (which includes tools and processes) that an organization requires to handle the large amounts of data and storage facilities. The term big data is believed to have originated with Web search companies who needed to query very large distributed aggregations of loosely-structured data. An Example of Big Data An example of big data might be petabytes (1,024 terabytes) or Exabyte (1,024 petabytes) of data consisting of billions to trillions of records of millions of people—all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured data that is often incomplete and inaccessible. Big Data and Types of Business Datasets When dealing with larger datasets, organizations face difficulties in being able to create, manipulate, and manage big data. Big data is particularly a problem in business analytics because standard tools and procedures are not designed to search and analyze massive datasets. This data, when captured, formatted, manipulated, stored, and analyzed can help a company to gain useful insight to increase revenues, get or retain customers, and improve operations.

B. Syferlock

SyferLock is an innovative provider of next-generation authentication and security solutions. SyferLock's patented software-based authentication and security solutions enable enterprises and government organizations to cost-effectively address strong authentication / 2 factor authentication to secure every access point, including computers, networks, online access and mobile devices, across a range of applications including proprietary networks, cloud

computing and mobile devices. SyferLock's user friendly software-based solutions deliver two-factor and multi-factor authentication through token less one-time passwords or PINs (OTPs) without the need for any additional hardware, tokens or client-side software, providing superior security along with greatly reduced Total Cost of Ownership (TCO). SyferLock's authentication solutions are available in enterprise and cloud editions. SyferLock's flexible methodology is easy to deploy, is extremely lightweight and can be deployed in a High-Availability (HA) cluster. Increasingly, enterprises are turning to SyferLock and its superior software-based two-factor and multi-factor authentication solutions to strengthen security, eliminate hardware tokens and to reduce TCO.

II. LITERATURE SURVEY

A.Bigtable: A Distributed Storage System

Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Bigtable, both in terms of data size (from URLs to web pages to satellite imagery) and latency requirements (from backend bulk processing to real-time data serving). Despite these varied demands, Bigtable has successfully provided a flexible, high-performance solution for all of these Google products. In this paper we describe the simple data model provided by Bigtable, which gives clients dynamic control over data layout and format, and we describe the design and implementation of Bigtable **Drawbacks**: This has not been implemented in real time application, and it is not sure how it handles tera bytes of data as HDFS is not used in this system.

B.The Next Frontier for Innovation

The amount of data in our world has been exploding, and analyzing large data sets—so-called big data—will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus, according to research by MGI and McKinsey's Business Technology Office. Leaders in every sector will have to grapple with the implications of big data, not just a few data-oriented managers. The increasing volume and detail of information captured by enterprises, the rise of multimedia, social media, and the Internet of Things will fuel exponential growth in data for the foreseeable future. **Drawbacks**: This project had not deals with security in team of real time application.

C.Big Data: How Do Your Data Grow

Data can be 'big' in different ways. National and international projects such as the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva in Switzerland, or the Large Synoptic Survey Telescope planned for northern Chile, are frequently cited for the way they will challenge the state of the art in computation, networking and data storage. **Drawbacks/Challenges**: No drawbacks, need to apply the concepts in real world and make it usable

D.Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications

In this method, to enhance the credibility of a composition plan, the evaluation of a service is promoted by some of its QoS history records, rather than its advertised QoS values. Besides, the k-means algorithm is introduced into our method as a data filtering tool to select representative history records. As a result, HireSome-II can protect cloud privacy, as a cloud is not required to unveil all its transaction records. Furthermore, it significantly reduces the time complexity of developing a cross-cloud service composition plan as only representative ones are recruited, which is demanded for big data processing. Simulation and analytical results demonstrate the validity of our method compared to a benchmark.**Drawbacks/Challenges**: No drawbacks, need to apply the concepts in real world and make it usable

E.Implementing Spam Detection Using Bayesian and Porter Stemmer Keyword Stripping Approaches

Unsolicited or spam emails are on the rise, where one's email storage inbox is bombarded with emails that make no sense at all. This creates excess usage of traffic bandwidth and results in unnecessary wastage of network resources. We wanted to test the Bayesian spam detection scheme with context matching that we had developed by implementing the keyword stripping using the Porter Stemmer algorithm. This could make the keyword search more efficient, as the root or stem word is only considered. Experimental results on two public spam corpuses are also discussed at the end. **Drawbacks/Challenges**: Key work search is not efficient

III. SYSTEM ANALYSIS

A.Existing System

Credit card usage and credit card transactions are increased rapidly. Existing system uses structured data, which is in SQL or oracle, analyzing structured data (10 TB) cost high for processor and transmitting data in internet is time consuming process.On the other hand Credit card fraudulent is increasing day by day, through there is OTP. Receiving OTP in mobile phone provides benefit, but in some cases, it will be a consuming more time and implementation cost and maintenance cost is also higher. It

is difficult to receive OTP if the user is in not reachable area, during this time it is difficult to access the account.

***B.Existing System Disadvantage*:**

- Attackers might hack/guess the pin
- Takes more processing time
- Implementaion cost is high
- CPU usage is high
- Uses only structured data
- Processing cost is high

IV. PROPOSED SYSTEM

Credit card transation are analysed using map reduce logic in HDFS, which is efficient way to analyze the data and respond quickly. This system uses MapReduce program to calculate the Input Keyword. MapReduce is a programming model for processing large data sets with a parallel, distributed algorithm on a cluster. A MapReduce program is composed of a Map () procedure that performs filtering and sorting and a Reduce () procedure that performs a summary operation.**Example**: Analyzing each credit card holder transatoion in different areas like shopping, grocery, hotels, petrol, this analysis will be useful for the bank to judge where customer is spending more amount , this will increase the revenue of the bank.Implemented **SyferLock** security for Credit and debit card transaction. There is no use in tracking the password because this password will be used only one, each transaction has different pin, it is not required to remember the pin each and every time. This system doesn't use additional OTP which user receives in mobile, instead it uses grid system that generate random password and displayed on the screen. User has to uses the random password based on their position to log in to the system.SyferLock concept is first time used in this Credit card management

A.Proposed System – Advantage

- Secured, Time consuming, Processing cost is low
- Applied Map reduce Concept – Big Data
- Unstructured data - HDFS
- Application oriented
- Easy analysis / reports for Business
- Reusable,Make Profit for business

B.Proposed System – Overcomes

- Shoulder Surfing,Screen rabbers/Recorders
- Video Recording, Keystroke Logging Malware or Hardware
- Man-In-The-Middle Threats, Browser Proxy Social Engineering, Phishing

A.SYSTEM ARCHITECTURE

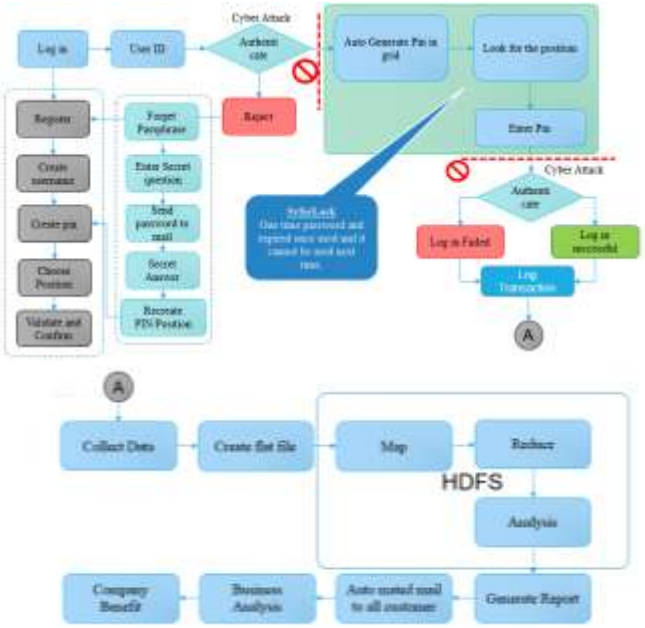

Figure 1 Architecture diagram

B.ER/DATA FLOW DIAGRAM

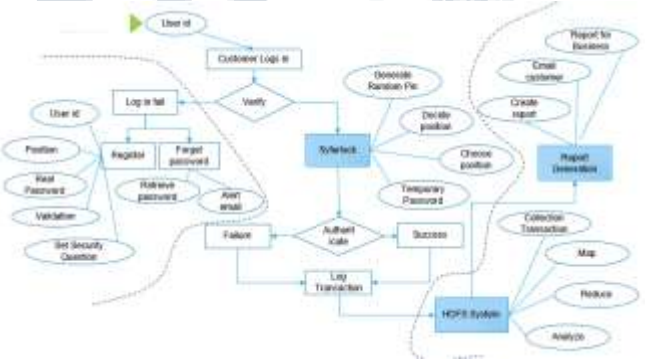

Figure 2 ER / Data Flow diagram diagram

C.USE CASE DIAGRAM

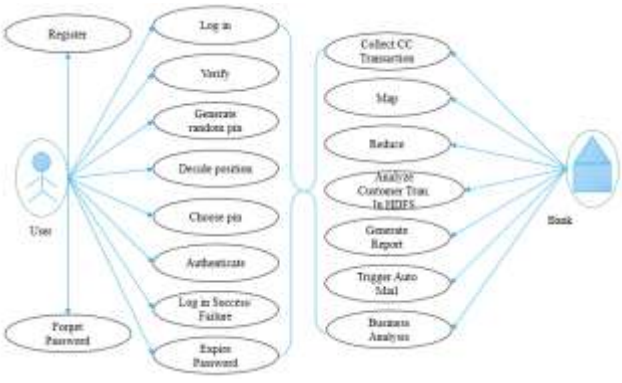

Figure 3 Use Case diagram

V. SYSTEM IMPLEMENTATION

I. Syferlock

SyferLock system provides patented authentication and security solutions. SyferLock delivers two-factor authentication solutions utilizing patented software-based grids to convert static passwords/PINs into secure one-time passwords/PINs (OTPs). SyferLock's software-based authentication solutions provide token-less OTPs, offering a simple, more secure way to access information while leveraging existing passwords and password infrastructure. SyferLock's flexible, adaptable solutions enable enterprises to cost-effectively address two-factor and multi-factor authentication across a range of use cases and with a range of platforms. SyferLock is market validated with a growing customer list and a number of awards from independent research firms and industry publications. SyferLock and its superior software-based authentication solutions to strengthen security, eliminate hardware tokens and to reduce Total Cost of Ownership (TCO).Key Features & Advantages of SyferLock's Authentication Solutions
Software-based/token-less OTP authentication

- Superior authentication and security
- Enterprise & Cloud editions
- Ease of deployment and use
- Greatly reduced TCO, both for direct and indirect costs (no tokens or token administration)

B.Grid View

SyferLock's unique methodology covers the authentication spectrum providing two-factor and multi-factor authentication utilizing patented software-based grids to convert static passwords/PINs into secure one-time passwords/PINs (OTPs) at each log-in without the need for any additional hardware, tokens or client-side software. SyferLock addresses the weaknesses of the traditional static password without the need for any additional hardware. SyferLock eliminates or mitigates a range of attacks like Key-Logging, Replay, Shoulder Surfing, Automated Attacks, Brute Force, Dictionary, Sniffing, Interception, Stored Browser Passwords, Cross Site Scripting, and Man-in-the-Middle

C.Position

User has to choose the position of the virtual password during register, which is applied during real time use.The Target Position is the location on the security pin pad that you will reference each time you login. The digit displayed in your selected location will be used to substitute for the corresponding digit in your secret 5-digit PIN to create a one time GridPIN

D.Radom number Generation

At log-in, a grid (as shown below) of cells is shown, each cell containing a static number or symbol in the center, and Random numbers in the corners that change with each authentication. User inputs the numbers corresponding to their pre-selected corner position in place of associated static password/PIN characters as their one-time password/PIN (OTP).

E.Password comparison

With a static PIN of "2490" and a pre-selected corner of "top left", the user would input a Grid PIN of "3347" for this log-in attempt.Upon every refresh and/or new log-in, the corner numbers randomly change, creating a new OTP. These single cells with number in the corners that change with every log-in are the foundation for SyferLock's patented software-based grids that are used to convert static passwords/PINs into secure one-time passwords/PINs (OTPs).

F.Secure log in

The password one used cannot be re used again, thus this system considered as a secured log in. And password stored in the database is encrypted. Cryptography algorithm is used to encrypt/decrypt the password

VI. BIG DATA

A.Loading data to Hadoop/Hive

Hive is a component of Hortonworks Data Platform(HDP). Hive provides a SQL-like interface to data stored in HDP. In the previous tutorial we used Pig which is a scripting language with a focus on dataflows. Hive provides a database query interface to Apache Hadoop.

B.Apply Map Reduce

MapReduce is a programming model for processing large data sets with a parallel, distributed algorithm on a cluster.A MapReduce program is composed of a Map () procedure that performs filtering and sorting (such as sorting students by first name into queues, one queue for each name) and a Reduce() procedure that performs a summary operation. The "MapReduce System" orchestrates by marshalling the distributed servers, running the various tasks in parallel, managing all communications and data transfers between the various parts of the system, and providing for redundancy and fault tolerance. Three steps in Mapreduce are Map Reduce and Finalize.**Key Points:** The SQL solution normally takes around100 seconds (just 1 sec for view creation and rest for grouping and joins. Using a temp table/indexes would speed this up).**Map Reduce will only take 6 seconds to execute.**There are other SQL and MapReduce solutions to this problem.

Generate charts/Reports

Using the result of Map reduce result reports are created and charts are generated. There are report generate can be used by individual user and business people. Reports are converted into HTML code and those will be mailed to individual customer and Business.

VII. COMPONENTS

A. Log In

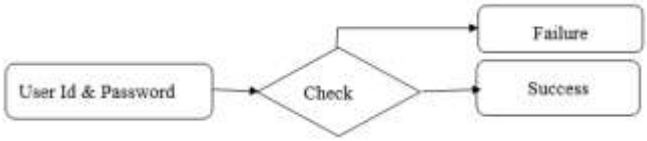

Figure 4 Log In

The application provides initional authentication for log in. User has to enter registered user id and password to log in, user who has not registered is not allowed to log in. System wont allow user if the user id and password in incorrect. Once user logged in successfully the session variable is maintained which can be used for validation throught the page. User can't access home page without logging, they will be redirected the log in page.

B. Register

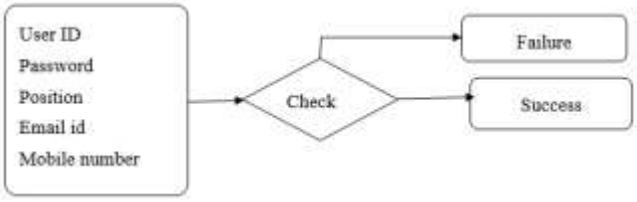

Figure 5 Register

New user has to register to log in to the system. Parameter included, user id, password, position, email id, mobile number. User has to provide unique userid else system won't allow to register. Password must be 6 digit numertic passoword. User has to choose the position in graphic interface UI and in drop down, both the position has to match. If all the validation succeeed then system allow to register the new user. If the user id is already exist the validation message is thrown to enter different unique user id. There is a separate method used to check if there is any duplicate user id has been entered. Bootstrap is used in register page makes screen responsive which is compatible in all device and all screen resolution.

C. Change / Forget Password

Figure 6Change Password

User is allowed to change the password by entering the coorect old password. New password allow only 6 digit numeric value. If incoorect old password is entered then the system won't allow to change the password. In this case user has to go for forget password option.

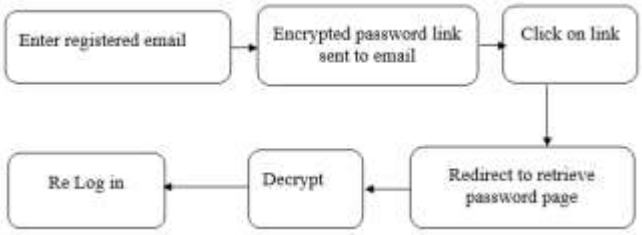

Figure 7 FORGET PASSWORD

If the user forget the password, this forget password link allow user to retrieve the password. Encrypted password is send to email with password retival link. Link from mail will navigate to password retrival page, and the password is decrypted and displayed to the user. **SendGrid SMPT** is servicing provider used to send email. Send Grid is can be used in any application to send mail.

GENERATE GRID PIN

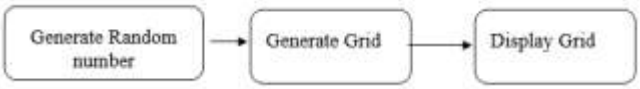

Figure 8Generate Grid Pin

Once the user logged in the grid view displayed with random pin. Using that user will view the random generated number with is used for login based on the position. There will be 9 grid displayed, each grid will be generating 8 random number around its position. Positions are Top, Bottom, Right, Left, Bottom Right, Bottom Left, Top Right, and Top Left. Based on registered password and positon user has to choose correct password to make transaction. Only 3 attempts are allowed to enter correct password, else it will be redireceted to the log in page.

D. Validation And Sql Connection

- Password should contain 6 digit numeric
- Password are encrypted
- Duplicate username are restricted
- Session state are maintained
- Random created are for single use
- Only encrypted password are send to registered email
- Default password is also randomly generated
- Highly responsive UI

SQL Database is used to save the data. User account data, transacation data are stored in SQL database. Table are structured and followed normal forms. Password saved the table are encrypted for a security purpose. No one can retrieve the password other than user.

E. Hadoop Implementation In Azure

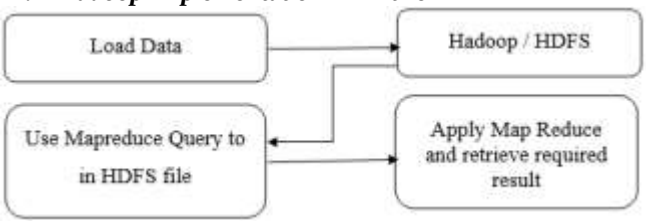

Figure 9Hadoop Implementation In Azure

Flat file is generated used as input for Mapreduce program. Flat file is initially loaded into HDFS for processing. Map reduce program counts the number of transaction caetogories and creates output file and placed in specified path in Azure HDFS which is used for report generation.

F. Mailing Reports

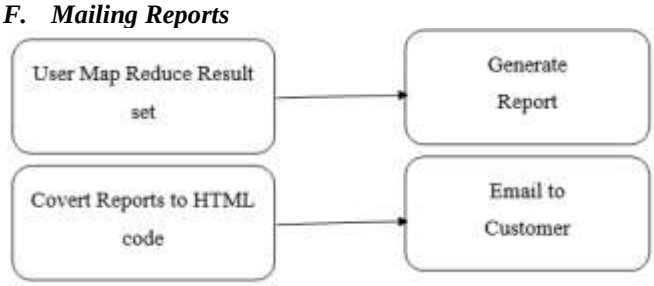

Figure 10Mailing Reports

Map reduce program is used in Bigdata HDFS to generate report. Input data is send to Azure HDFS which has map reduce program and sends input to generate report. Send Grid is email service provider which is used to send email from web application. Application sends HTML code, from mail address, to mail address, SMTP user id password to SendGrid service as a parameter which will send email to concern user.

VIII. CONCLUSION

This project provides secure log in without any external OTP generation, as this uses pattern logic with random grid PIN. Also the transaction log that logged in Credit and Debit card transaction is analyzed using Microsoft Azure HDFS. As HDFS uses Map Reduce logic, analyzing Big Data for Credit and Debit card is easier, which is help full for the Business to grow better.

APPENDIX-SCREENSHOT

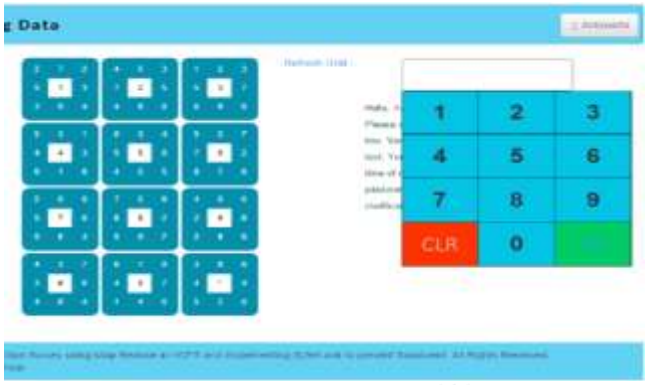

Figure 11. Grid view sample

ACKNOWLEDGEMENT

I give all glory and thanks to our God almighty for showering upon, the necessary insight and grace for accomplishing this project work.

I am in debated to our honorable Chairman and Founder of Dhanalakshmi Srinivasan group of Institutions, *shri.A.Srinivasan*, for his kind encouragement. I am very much happy to thank our beloved Principal, *Dr.T.Elango, M.E., Ph.D.*,for his unflinchingdevotion andcontinuous encouragement motivated me to complete this project work phase I.

I am thankful to our Vice Principle and Head of the Department, *Prof.S.Nandhakumar, M.Tech, (Ph.D)*, who encouraged me a lot to complete my project work phase I successfully.

I am greatly thankful to the project coordinator *Mrs.J.Sasidevi, M.E., (Ph.D.)*, and my internal guide *MR.V.Gokulakrishnan, M.E., MBA.* Assisant Professor.

REFERENCES

[1] Big Data and IT-Enabled Services Ecosystem and Coevolution

[2] C. Lynch, “Big Data: How Do Your Data Grow?” Nature, vol. 455, no. 7209, pp. 28-29, 2008.

[3] F. Chang, J. Dean, S. Ghemawat, and W.C. Hsieh, “Bigtable: A Distributed Storage System for Structured Data,” ACM Trans.Computer Systems, vol. 26, no. 2, article 4, 2008.

[4] G. Linden, B. Smith, and J. York, “Amazon.com Recommendations

[5] https://azure.microsoft.com/en-us/documentation/articles/storage-monitor-storage-account/

[6] http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4597240&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4597240

[7] Item-to-Item Collaborative Filtering," IEEE Internet Computing,vol. 7, no. 1, pp. 76-80, Jan. 2003.

[8] i2MapReduce: Incremental Map Reduce for Mining Evolving Big Data

[9] Internet of Things and the Credit Card Market

[10] J. Manyika et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity," 2011.

[11] KASR: A Keyword-Aware Service Recommendation Method on Map Reduce for Big Data Applications

[12] "The Internet of Things", ITU Internet Reports, 2005

[13] Towards a Restful,"Plug and Play Experience in the Web of Things".

[14] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and Evaluating Choices in a Virtual Community of Use," Proc.SIGCHI Conf. Human Factors in Computing System (CHI '95),

[15] W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications," IEEE Trans. Parallel and Distributed Systems, 2013.

[16] Y. Chen, A. Cheng, and W. Hsu, "Travel Recommendation byMining People Attributes and Travel Group Types from Community-Contributed Photos," IEEE Trans. Multimedia, vol. 25, no. 6,pp. 1283-1295, Oct. 2013.

[17] Z. Zheng, X. Wu, Y. Zhang, M. Lyu, and J. Wang, "QoS Ranking Prediction for Cloud Services," IEEE Trans. Parallel and DistributedSystems, vol. 24, no. 6, pp. 1213-1222, June 2013.

AUTHORS

First Author – R.Anbuvizhi, ME-CSE in Dhanalakshmi Srinivasan Engineering College,
Perambalur- 621212, TamilNadu, India.
Email Address: r.anbuvizhi@gmail.com

Second Author – V.Balakumar, Completed BE-CSE in Dhanalakshmi Srinivasan Engineering College,
Perambalur- 621212, TamilNadu, India @ 2011.
Working as System Engineer in Tata Consutancy Service, Chennai, since Feb 2012.
Email Address:bala101289@gmail.com