Abstract

KYC (Know Your Customer) is becoming a critical gatekeeper process for financial institutions, the world over, to safeguard against financial frauds, terrorist funding and money laundering. It involves collecting basic identity & address information about the customer. Regulatory agencies have been coming down heavily on defaulting organizations thereby forcing many of them to invest in state of the art financial transaction surveillance systems. One of the biggest challenges that the industry faces while it steps up monitoring is the sheer size of the data, speed of generation of this data and complexity arising out of multiple & non-standard formats. World continues to generate data at an unprecedented pace. This generation calls it Big Data. Traditional data warehousing techniques, that have stood the test of time, have lately failed to deal with volume, velocity & variety of data (Big Data). This is where Big Data Analytics has emerged victorious. Rather than relying on structured data techniques, Big Data analytics attempts to rely on basic techniques like pattern matching, divide & conquer & decentralized processing to solve real life problems. Though this technology is still new but it has already shown signs of maturity. This paper attempts to study Know Your Customer process, articulate the challenges involved and highlights the short-comings that the systems today have in effectively implementing KYC guidelines (especially in large Indian banks). It then, using real life examples, presents a credible solution using Big Data Analytic techniques like Fuzzy Matching
& MapReduce. Authors are confident that the framework of the solution that has been provided can lead to a working prototype in a short span of time.

References

- Improving the Integrity of Identity Data - Data Matching Better Practice Guidelines, 2009
- Gianmarco De Francisci Morales, Aristides Gionis, Mauro Sozio, “Social content matching in MapReduce”; research paper presented at 37th International conference on Very Large Databases (VLDB), 2011

Index Terms

Computer Science

Information Sciences

Keywords

KYC   Aadhaar   Unique Identity Proof   Fuzzy matching   Identity theft   MapReduce