



CASE STUDY

Putting our expertise in automation and validation together to create a valuable AML solution for a UK-based bank



Infosys
be more

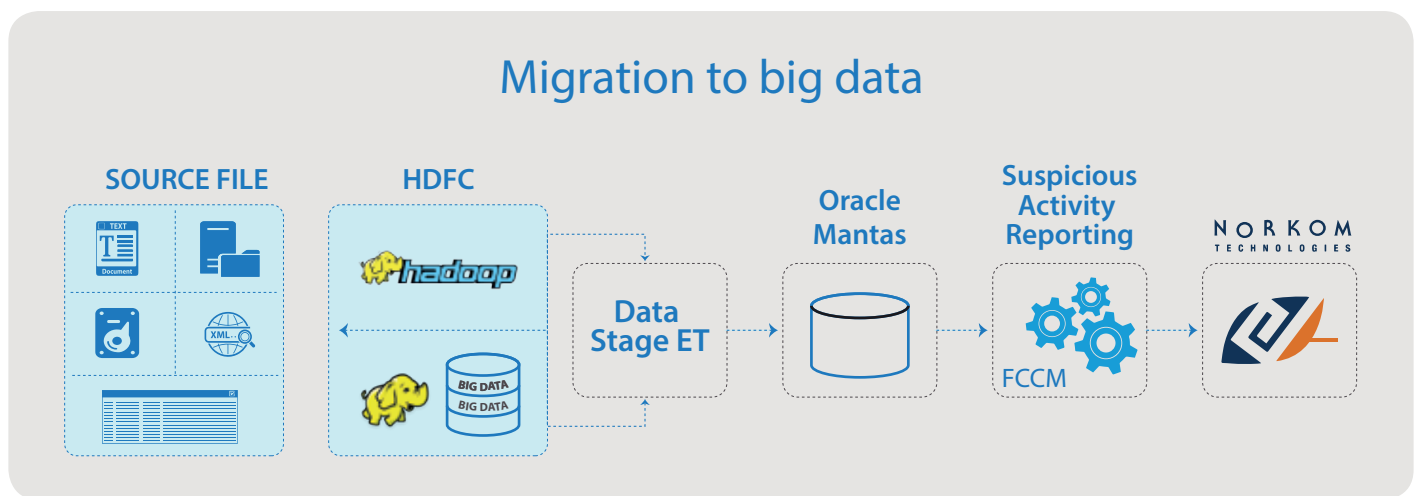
About the client and Infosys partnership

The client is a leading UK-based bank with a global presence in retail, investment, and commercial banking. It has around 6,600 offices in 80 countries across Africa, Asia, Oceania, Europe, and America to serve their customer base of 60 million. In order to operate seamlessly across the globe, the client implemented an Anti-Money Laundering (AML) program to detect unusual and possibly illegal transactions. This solution was implemented for core banking functionalities and was a Norkom product. However, as the client required a specific AML solution that also processed data from affiliates and correspondent banks to overcome the shortcomings of the Norkom-based product, they entered into a partnership with Infosys in October 2014.



The challenges

Migration to big data



The client faced the following five primary challenges:

- Huge amounts of masked data needed validation — data from the source systems were masked and it was essential that the validation of masking logic was completed (to prove data integrity) before the start of actual validation. This process took eight days for 32 source files
- Complex staging, ETL (extract, transform, and load), and Hadoop transformations — it was critical to maintain traceability and transformation logic from the source system to Oracle Mantas to achieve 100 percent test coverage
- Absence of a Graphical User Interface (GUI) to validate Hadoop / Hive data — GUI viewers were not allowed to validate Hadoop data due to security reasons, making comparison difficult
- Different formats of data in relational database management system (RDBMS) and Hadoop — Hadoop had a file-based storage whereas the source system had relational storage of data. There was no uniform method to compare the output from the source file to the Hadoop / Hive output
- Lack of end-to-end (E2E) dynamic and heterogeneous systems — the source system and staging area were located in one UNIX box and the Hadoop Distributed File System (HDFS) was housed in another, separate UNIX box. At the same time, Oracle Database 11g was housed in Windows. As a result, an end-to-end master script was required to orchestrate the validation

The Infosys solution

Infosys introduced many innovations in this project like Hive Output Formatter (HOF) and Transformation Traceability Matrix (TTM). We pioneered an end-to-end testing solution for Anti-Money Laundering (AML). The following are the primary components of our solution:

- **Automated utility**
Java-based, masked data validator compares masked and unmasked data with the data dictionary. A detailed result output is generated comprising of unmasked, masked, and mismatched data
- **Comprehensive TTM**
A single source for initial and target transformation. It links the source (database objects, files) with the target (database objects, files), and also showcases impacted measures (facts) and dimensions for every transformation rule. This helps in creating test scenarios for specific functional changes as well as overall impacted areas
- **Query output formatter**
Introduced a Hive Output Formatter to format and analyze Hive query outputs as well as to create a pipe delimited file output at specific locations for validation
- **Text-to-text comparator**
Java utility created to compare two pipe delimited text files, field by field. JavaScript can be used for text files of any size and comparisons are done based on the metadata file
- **End-to-end regression automation framework (in HDFS / Hive and ETL)**
Triggered through a remote UNIX script to ensure one-time synchronization of all UNIX boxes using Rivest-Shamir-Adleman (RSA) for persistent connectivity



End-to-end big data testing combined with accelerators

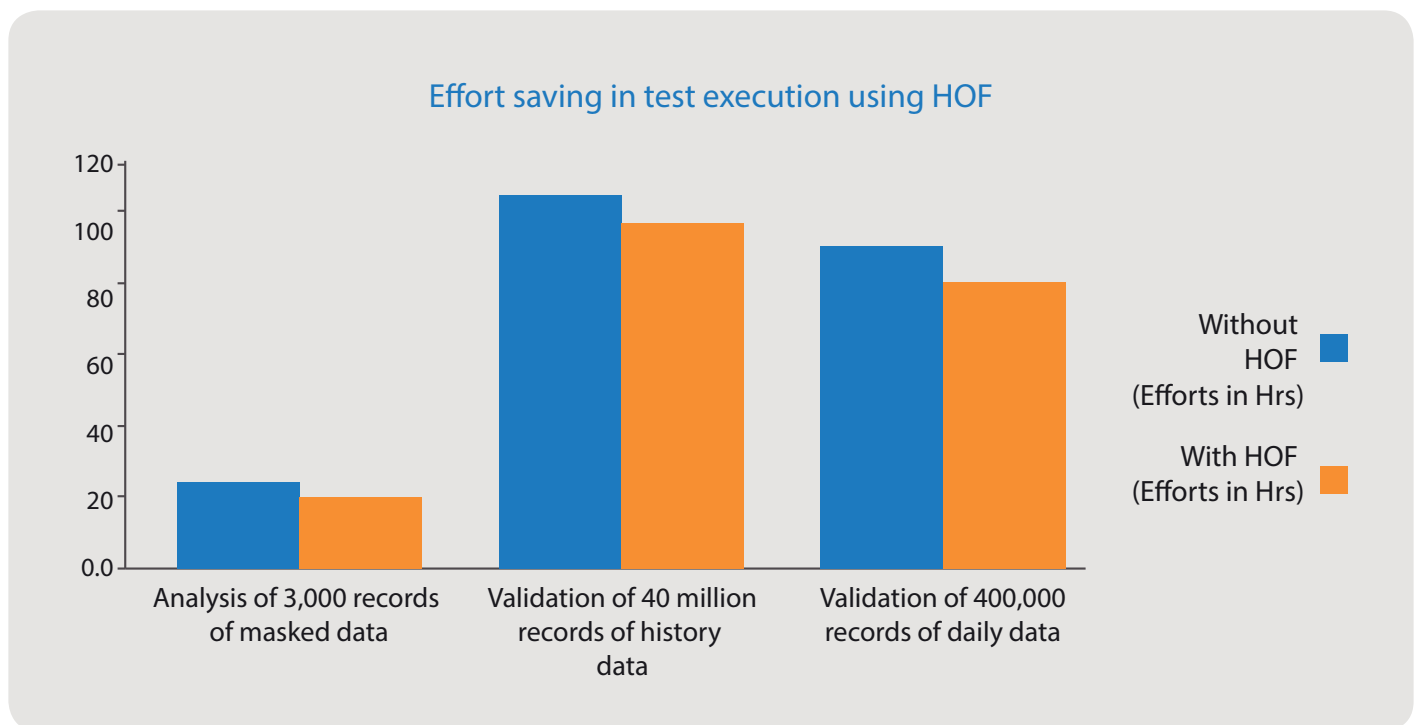
The partnership has been growing strongly after the successful testing of our AML solution through validation of:

- Filtering criteria by writing specific queries in the Hadoop data lake (incorporating filter criteria in Hive queries) and comparing it with the source data (files)
- Data in the Hadoop data lake, after transformation, by using custom Pig Latin scripts to check if the records are transformed according to the requirements
- Specific data being extracted from IBM InfoSphere DataStage 9.1 ETL tool by comparing the query output against the output provided by ETL tool jobs
- Data moved into the Suspicious Activity Report (SAR) by comparing it with the query output of both Oracle MANTAS and NORKOM
- Reconciliations between Oracle MANTAS and Norkom to ensure that all data has been transferred without errors

Business benefits

- 50 percent effort reduction in masking validation by using a Java-based masking utility
- 15 percent effort reduction in requirement analysis by using the TTM
- 20 percent reduction in Hadoop execution effort by using the Hive output formatter
- 50 percent effort saved by using text-to-text comparator (RDBMS to Hadoop comparator)
- 15 percent effort saved in overall test execution cycle by using E2E regression automation
- 40 million history records and 140,000 daily transactions are handled by the solution newly developed by Infosys
- Four utilities reused across big data testing programs at Infosys, yielding a cumulative productivity improvement of 30–35%
- Reusable test strategy for big data migration testing, thereby cutting down the test planning time by 20–25%

Tangible efficiency improvements through automation by using HOF (Hive Output Formatter) and Text-to-Text Comparator



For more information, contact askus@infosys.com

Infosys
be more

© 2017 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.