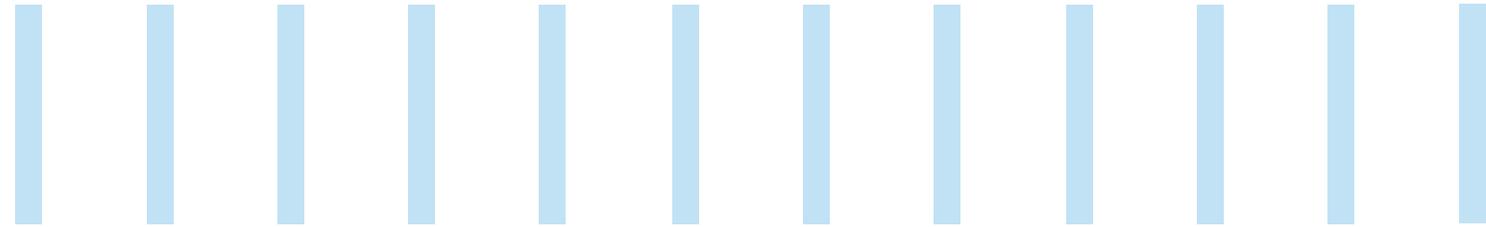




# AUTOMATING BIG DATA TESTING FOR PERFORMANCE



Just as the number of applications and their interconnectivity is growing, so is the amount of data being collected from customers. This explosion of data is a boon for organizations who are eager to use it to gain insights into customers' needs and expectations.

To harness the available information properly, organizations need to build capabilities to manage the increased volume, variety, and velocity of data; to analyze this new complexity in real-time; to consider the appropriate architecture required to ensure all data types are supported; to **integrate data in the broader Business Intelligence (BI) strategy**; and most importantly, to have a positive business impact on the organization.

# ACTIONABLE INSIGHTS FROM BIG DATA

These requirements are well known to the technology community. In a recent study by Greyhound Research, a leading global analyst firm, 87% respondents said they were unsure of their approach to tackling big data, and 62% said they were grappling with issues around business processes, methodologies, governance, and security. The same study also highlights another critical aspect – 90% respondents confirmed cloud computing, the use of which is a given in big data projects, further complicates the integration, management, and validation of big data.

For organizations trying to ascertain the best approach to managing big data, a good starting point is to **identify the possible points of failure** and define a testing strategy accordingly. Furthermore, organizations must set up an **optimal test environment** and support it with a comprehensive and preferably **automated validation framework** in order to certify that the data is as per their business requirement.

A leading US-based **jewelry retailer**, an Infosys client, wanted to gain actionable insights from their data but was struggling with slow performance of on-premise applications and high infrastructure costs. Infosys took up the **migration of data** for the client from big data systems like Apache Hive and Hadoop to **Amazon Web Services (AWS) big data platform** along with the validation of reports generated from this migrated data.

Infosys helps clients across industries facing similar data challenges to achieve the benefits of quality, time-to-market, and costs. Infosys' big data and analytics validation offerings cover the entire big data journey, not only automating data quality monitoring, but validating data movement from multiple sources and ensuring data integrity, helping clients fine-tune **pattern-matching algorithms** and increase the **usefulness of unstructured data**.

# END-TO-END AUTOMATED VALIDATION

For the retailer, which had huge and varied data sets coming from different sources such as enterprise data warehouse sales mart, enterprise customer master data and third- party sources such as Adobe campaign data, a huge amount of data had to be migrated to AWS. The data also had to be validated to ensure that it was accurate, complete, and free of anomalies, and this had to be done at multiple stages: as the data moved from on-premise systems to Amazon S3, then to the big data platforms, from there to the Redshift warehousing solution and finally to the Tableau report.

Because of the huge amount of data involved, and the different technologies and platforms involved, the solution **called for automation**. For this, we used a proprietary Infosys validation tool – **Infosys Data Testing Workbench (IDTW)**. Typically used for on-premise validation, Infosys customized IDTW to cater to the project's requirement of offering support for cloud platforms as well. We used the tool for **end-to-end automated data validation** for migration of on-premise data to Amazon cloud. We also created **repeatable regression test suites** for multiple test executions seamlessly.

Automated data validation through our proprietary IDTW tool helped **save 40% of effort that would have been required to validate manually**. This also gave the client a single automation tool to support disparate data sources. The IDTW tool also helped the client achieve the performance benchmark of **50,000 records validation in less than 5 milliseconds**, and helped them gather better insights from customer data at greatly reduced costs.



# AUTOMATING BIG DATA TESTING FOR PERFORMANCE: THE FIVE TAKEAWAYS

- 1 **Migrate** big data systems from on-premise applications to cloud to improve performance and reduce high infrastructure costs
- 2 **Validate** data through different stages of data journey to maintain data integrity
- 3 **Automate** to validate data at scale
- 4 **Create** repeatable regression test suites
- 5 **Expect** new challenges when using cloud

# BIG LEARNING:

As organizations increasingly hinge their business models on data, the need for them to plan before execution and be precise and cautious during execution cannot be overemphasized. The need to follow this well-thought-out approach is all the more critical when organizations are juggling multiple projects like migration to cloud along with managing new sources and types of data.

## WE DID THIS FOR THEM. WE CAN DO IT FOR YOU.

To learn more about validation solutions for big data, reach out to us at [askus@infosys.com](mailto:askus@infosys.com)

For more information, contact [askus@infosys.com](mailto:askus@infosys.com)

**Infosys**<sup>®</sup>  
Navigate your next

© 2018 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.

[Infosys.com](http://Infosys.com) | NYSE: INFY

Stay Connected     