

INFOSYS DATA WORKBENCH

Infosys[®]
Navigate your next

Overview

While most organizations understand the importance of data, ensuring quality is always a challenge.

For data to be useful, it must be of a high standard, be consistent and unambiguous. Data also continues to be a critical element as businesses look to leverage it in new ways. With the shift towards real time

analytics & insights, it has become more important than ever to have high quality data always available.

There are several reasons for poor data quality. As businesses become more complex, data becomes more distributed and disparate. Cloud adoption has created its own set of challenges like database

merges due to cloud integration processes. Organizations also find that data is stored in non-standard formats, is incomplete or incorrect, or duplicated.

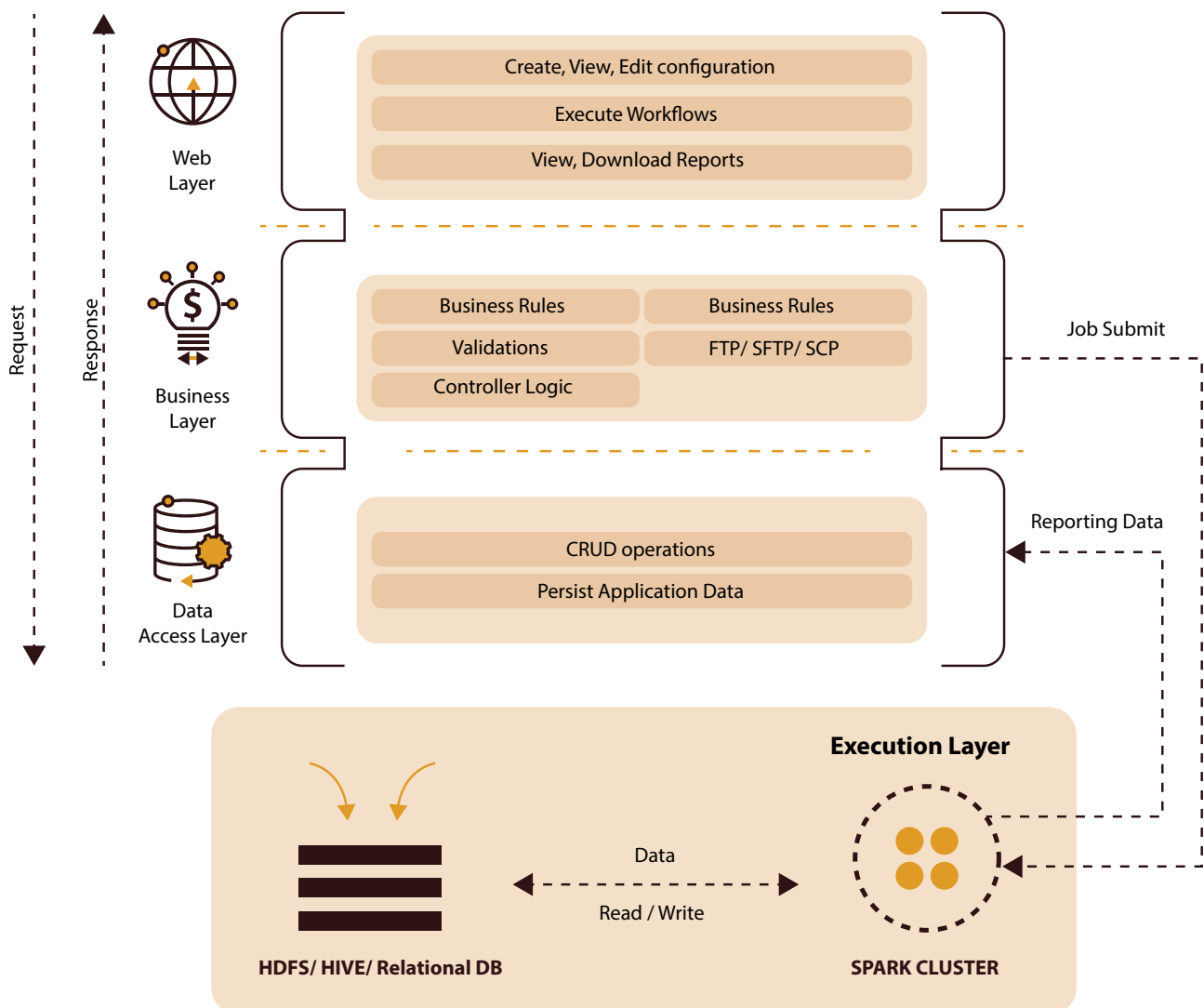
The above factors throw up the need for a robust Data Quality Solution that is cost efficient and provides faster results.

Infosys Data Workbench

Infosys Data Workbench (iDW) is a highly scalable Data Quality Solution coupled with light-weight analytical MDM capabilities. It is a single tool that can perform varied operations on an

organization's data. iDW comes with an AI-powered core that helps perform large-scale data quality operations with just a few clicks. A powerful Apache Spark driven core also helps perform high speed

analysis for data at scale. Infosys Data Workbench is designed to be easy to use and can be leveraged to gain insights into data, without the need for deep technical knowledge.

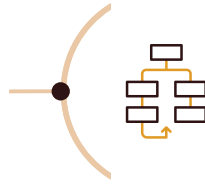




Benefits

Latest & Hybrid Algorithms

The latest developments in deterministic, fuzzy, phonetic techniques have been incorporated.

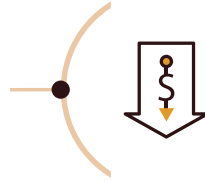


Cloud Ready Deployment

Web-based interface and open source stack enables cloud deployment at low cost.

Low Implementation & Operating Cost

Runs on commodity hardware and can scale up based on requirements. Is also flexible and keeps operating costs in check with a low licensing fee.

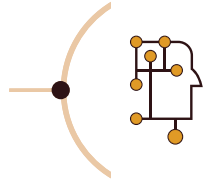


Data Sources Support

Supports multiple input & output formats and data sources.

Automated & Script Based Setup

Automated scripts for tool installation including cluster setup across multiple platforms.

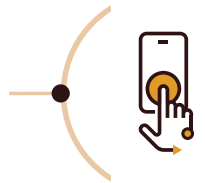


Better Security

LDAP authentication, SSO and role-based authorization enabled.

Simple, easy to understand and intuitive GUI

Anybody can start using the tool with just a couple of hours of training and no dependency on tool experts.

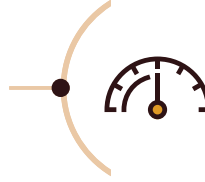


Scheduling Options

Options for scheduling jobs multiple modes – one time / recurring.

Speedier Results

Apache Spark based execution engine provides fast results.



Highly Scalable Solution

Spark based execution engine to help scale horizontally.

Rich & Insightful Reporting

Reports are available in dashboards as well as in downloadable formats – PDF & Excel.



Machine Learning Based Enhancements

Machine learning based models for outlier detection, classification, enrichment.



Use cases

Data Quality

Data Profiling:

Assess the quality of data and identify the data outliers using out of box or pre-configured rules.

Data Cleaning & Standardization:

Cleanse and standardizes the data using out of box rules. Address data cleansing using 3rd party APIs.

Business Rules Validation:

Validate data using complex business rules and monitor for threshold breaches.

ML based Outlier Detection:

Identify probable outliers in the data without the user specifying any rules by using unsupervised learning algorithm.

ML based Data Enrichment:

Populates values for business-critical attributes by using ML models generated on historical data available in the system.

Analytical MDM

De-duplication:

Identify duplicate records in the source system by using algorithmic data matching techniques.

Consolidation/Golden Record Generation:

Data from various systems can be combined and duplicates are removed with data stewardship to approve or reject, marginally failed records.

Data Enrichment:

Enrich data by supplementing data from trusted sources.

Linking/Grouping:

Create links between multiple datasets (Internal/External) using data matching techniques

Discovery

Data Discovery:

Identify the sensitive and non-sensitive data elements present in a data source. Classify, categorize and label the data according to the business needs. Analysis can be performed on both structured and unstructured data.

Rule Miner:

Understand the relationship between various business critical attributes and generate various rules for them along with a list of outliers.

DATA QUALITY



- Data Profiling
- Data Cleaning & Standardization
- Business Rules validation
- ML based Outlier Detection
- ML based Data Enrichment

ANALYTICAL MDM



- De-duplication
- Consolidation/- Golden Record Generation
- Data Enrichment
- Linking/Grouping

DISCOVERY



- Data Discovery
- Rule Miner





Client Stories



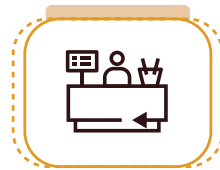
A Large European Insurance Company:

iDW analysed their data quality issues, performed de-duplication using Fuzzy Match Techniques and One-time Data Cleansing. The cleansed data reduced processing times leading to a 45% increase in productivity due to reduced cost of quality.



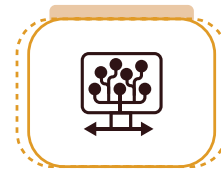
An International Registrar & Classification Society:

iDW was leveraged to perform Identity Resolution and enrich their master-data with External Directory Data and for Golden Record Generation of Customer Data. The enriched data helped the organization offer customized campaigns to its customers. The automated data enrichment helped eliminate manual processes and resulted in an overall productivity improvement by 35%.



A Large US Retailer:

Extracted data from source files, performed data quality analysis and suggested data cleansing rules. De-duplication analysis was performed using algorithms via Fuzzy Match Techniques. This resulted in a 40% increase in productivity due to reduced cost of quality.



A Large US Technology Company:

Enabled ML based predictive model algorithms were used to classify unverified data listing and also pattern identification using clustering. The predictive models helped improve productivity by 40%.

About Us

The incubation center of Infosys called 'Infosys Center for Emerging Technology Solutions' (ICETS) focuses on incubation of NextGen services and offerings by identifying and building technology capabilities to accelerate innovation. The current areas of incubation include AI & ML, Blockchain, Computer Vision, Conversational interfaces, AR-VR, Deep Learning, Advanced analytics using video, speech, text and much more.

To know more, please reach out to ICETS@infosys.com

For more information, contact askus@infosys.com



© 2020 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.