Infosys Data Testing Workbench (IDTW)

Next Generation Data Validation Solution

Infosys®
Navigate your next

Data validation services - Infosys Point of View

Infosys® | Center for Emerging Technology Solutions

The demand now is for interactive analytics with multi-channel support and scalability . There is huge scope for improvement and enhancements in this space by offering a simple end-to-end solution for analytics testing.

	Data Stores	Data Processing	Validation	Insights
Traditional	Schema Rigid RDBMS Structured	ETL Jobs Data Migration File/RDBMS to RDBMS/file	Data Comparison Table Analysis Report Testing Test Management	Reporting - Counts, Max, Min, Avg, Data Pre-processing SQL Queries Summary

Drivers	4Vs	Emerging Tech	Agile/ DevOps	Data science	Interactive	Open Stack

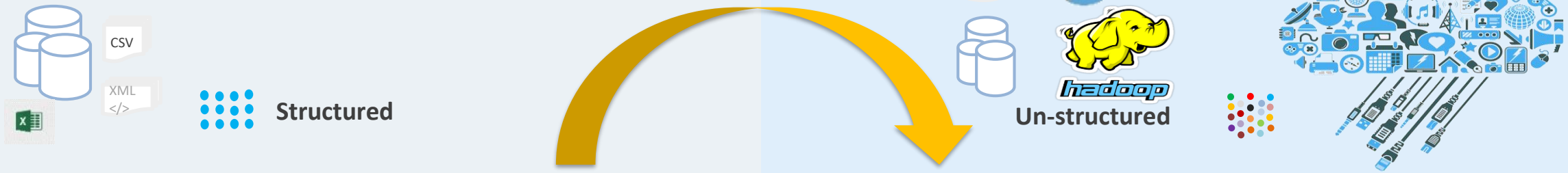
Variety, Volume, Veracity, Velocity

Emerging Scope	Schema-less Unstructured Image/Video/Social Media Huge Volume – TBs, ZBs	Multi-source, omni-channels Lambda-Arch, Data Lakes, Cloud DBs, IoT Data Pipelines	ALM, DevOps, CI/CD Pipeline Validation Data Lakes Health AI-based Detection	Real-time Analysis, Test Interactive Visualization Insights anywhere	AI/ML on data Collaboration Open Standards	

Infosys® Navigate your next

Data Testing in today's context

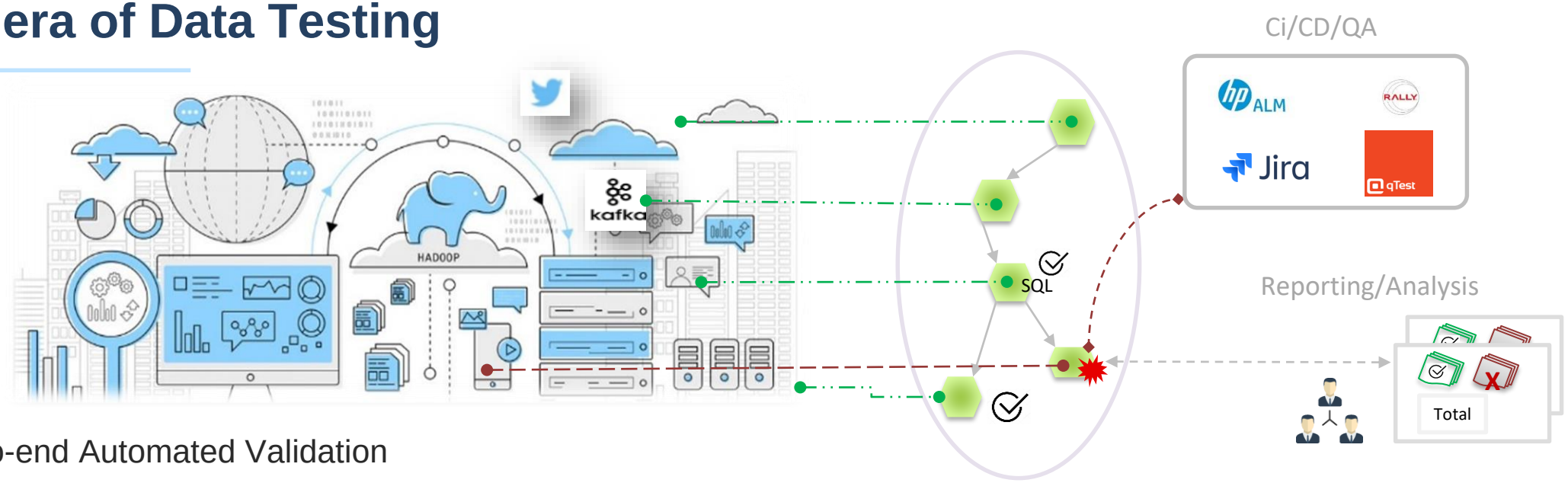
With the rapid rise in data stores, streaming apps, Cloud & IoT – the expectation is to automate right from *event* to *access* time while having the capability to support data both in motion and at rest. Enterprises in today's setting need to be enabled with simple, scalable end-to-end validation of streaming systems and multi-channel scenarios with seamless integration with CI/CD QA tools.



Data stores	Data processing	Validation
Schema Rigid	ETL Jobs	Data Comparison
RDBMS	Data Migration	Table Analysis
Structured	File/RDBMS to RDBMS/file	Report Testing
		Test Management

Multi-source, Omni-channels	**Real-time analysis, test**	**Pipeline Validation**
AI / ML on data	**Interactive Visualization**	**Data Lakes Health**
Open Standards	**End-to-end validation**	**Insights anywhere**

IDTW – A comprehensive one-stop solution for this new era of Data Testing



End-to-end Automated Validation

- Supports data in rest, data in motion
- Multi-protocol Support
- Data migration Validation – On-prem to Cloud DBS, AWS, Azure, etc.
- Data Quality for all heterogenous files, RDBMS, Big Data, Cloud Data Stores

- Test Orchestration Pipeline – catering to streaming, multi-interface data validation with visual monitoring
  -Pre-built automated QA templates
- SQL for whole no-SQL world
- Interactive Notebook-based Visualizations

Enterprise Features

QA Workflow, Test Governance, Management, Audit-trail, Security, Reporting, Batch Scheduling

OAuth2
LDAP/Custom
Security

Install

Micro-services Platform

Cloud Native

- **Single node**
- **Cluster to N laptops/desktops**

Scalable

Infosys Confidential

© 2020 Infosys Ltd.

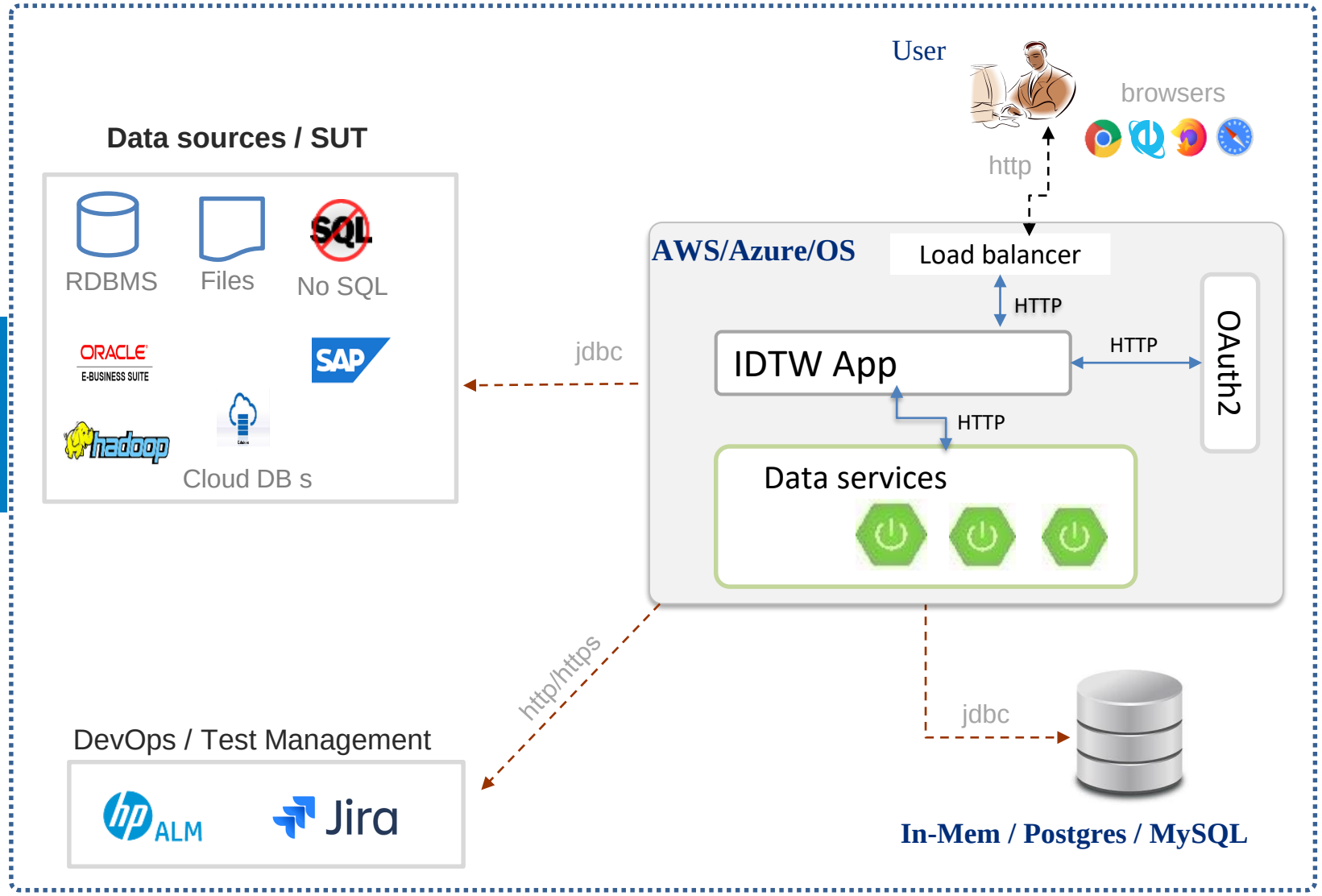
IDTW has helped its users realize benefits on multiple aspects across verticals

Infosys® | Center for Emerging Technology Solutions

Vertical	Challenge	Benefit
Banking	Client has Global Data Warehouse (GDW) & a 3 Terabyte Data Warehouse that connects to 120+ source systems & 100+ downstream systems	» Cycle Time Reduction by 20-22% » 90% reduction in Data Validation
Banking	Client Group integrated data warehouses of 4 banks where the technology landscape was highly complex	» COQ reduced to 15.9% » ZD - 588K AUD Approved
Insurance	Client enhanced their complex agency compensation repository for new metrics calculation	» Cycle Time Reduction by 15% » 70% Effort Savings
Telecom	Project involved testing across various data sources with huge volumes of data	» 100% Test Data Coverage
Retail	Client planned to optimize and migrate their BW landscape to HANA Landscape within 6 months	» Cycle Time Reduction by 40% » 60% Effort Savings

Infosys®
Navigate your next

IDTW architecture and deployment in the eco-system



Where all can we deploy / setup ?

On-Premise

- Laptop/Desktop
- Servers/VMs

- Suitable for on-premise setup typical desktop/laptop/VMs
- Most of the setups in implementations use this

Cloud

- Can be hosted on AWS,Azure easily
- On cloud installations reach the tools team for choosing apt instances(*t1,t2,t3 etc*) in Cloud

docker

- Suitable for running in container based scenarios – internal cloud, hybrid, GCP platform
- Image can be hosted on any environment : Linux/ windows/ cloud

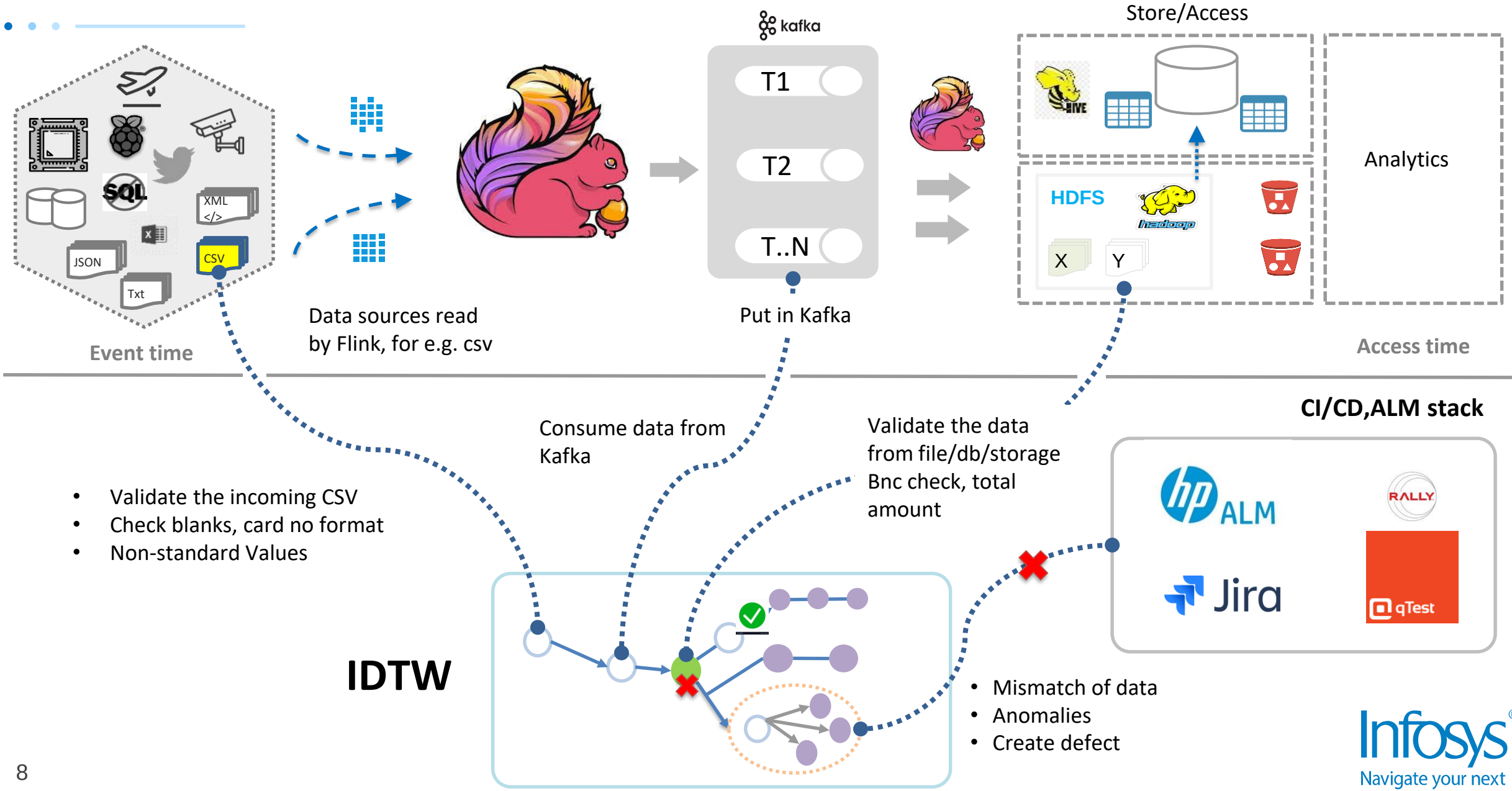
Pre-requisites

- Windows/Linux (win 10,server, red hat linux) with Java 8 installed
- PostGres/MySQL/In–mem

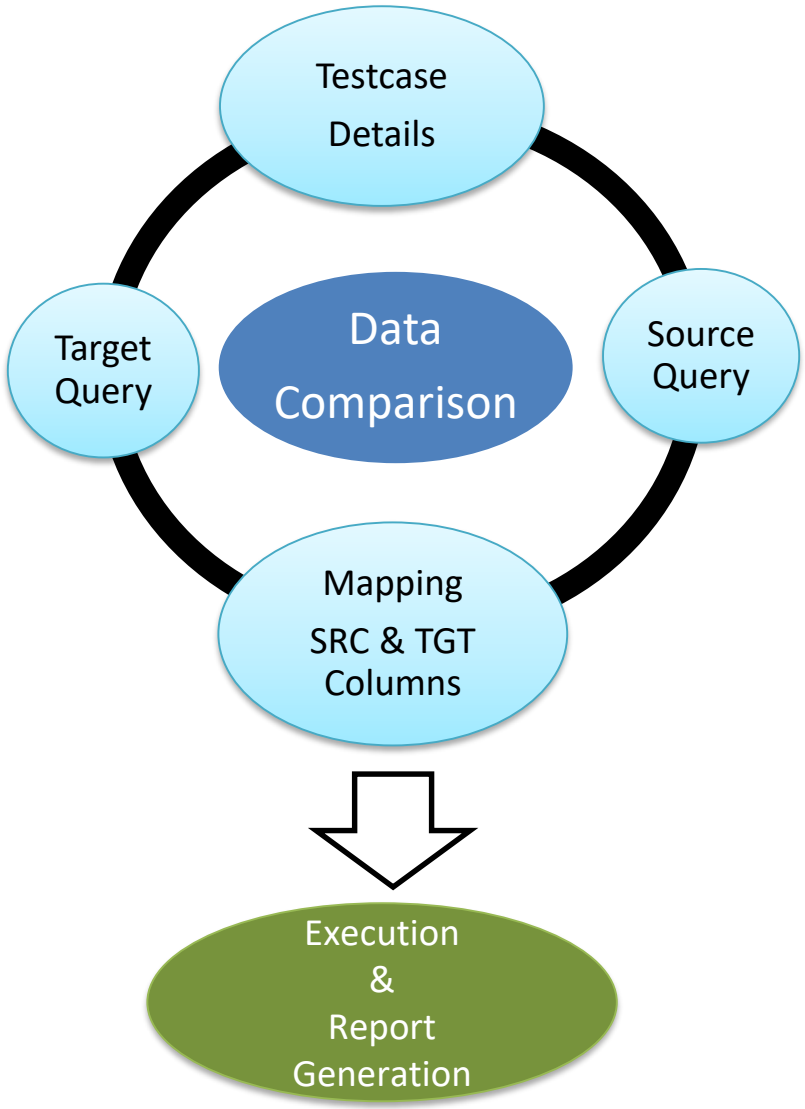
Note: postgres/mysql if user work need to persist

Infosys®
Navigate your next

Test pipelines – streaming scenarios



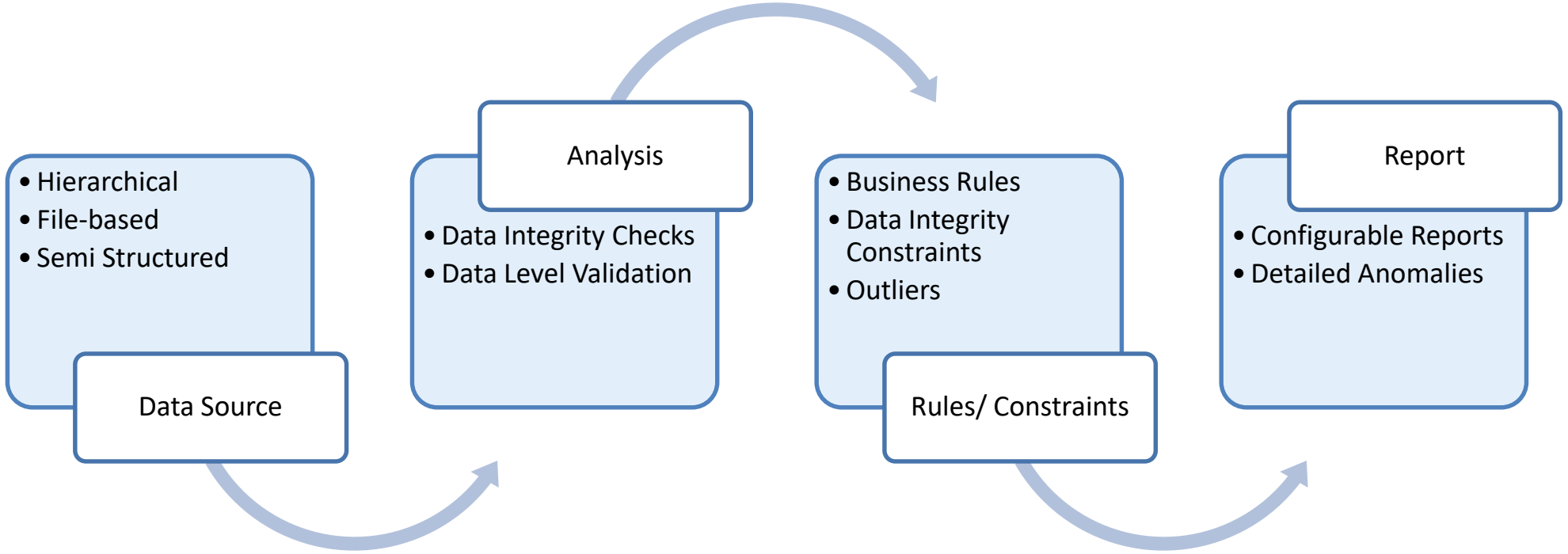
Data Comparison Module



- Data Comparison between any source Vs. any target
- Field Level Comparison
- Detailed Anomaly Reporting for easy defect fixing

Infosys® Navigate your next

Data Quality Checks on Big Data



Metadata
- Constraints (PK, FK)
- column Data Type
- Nullability

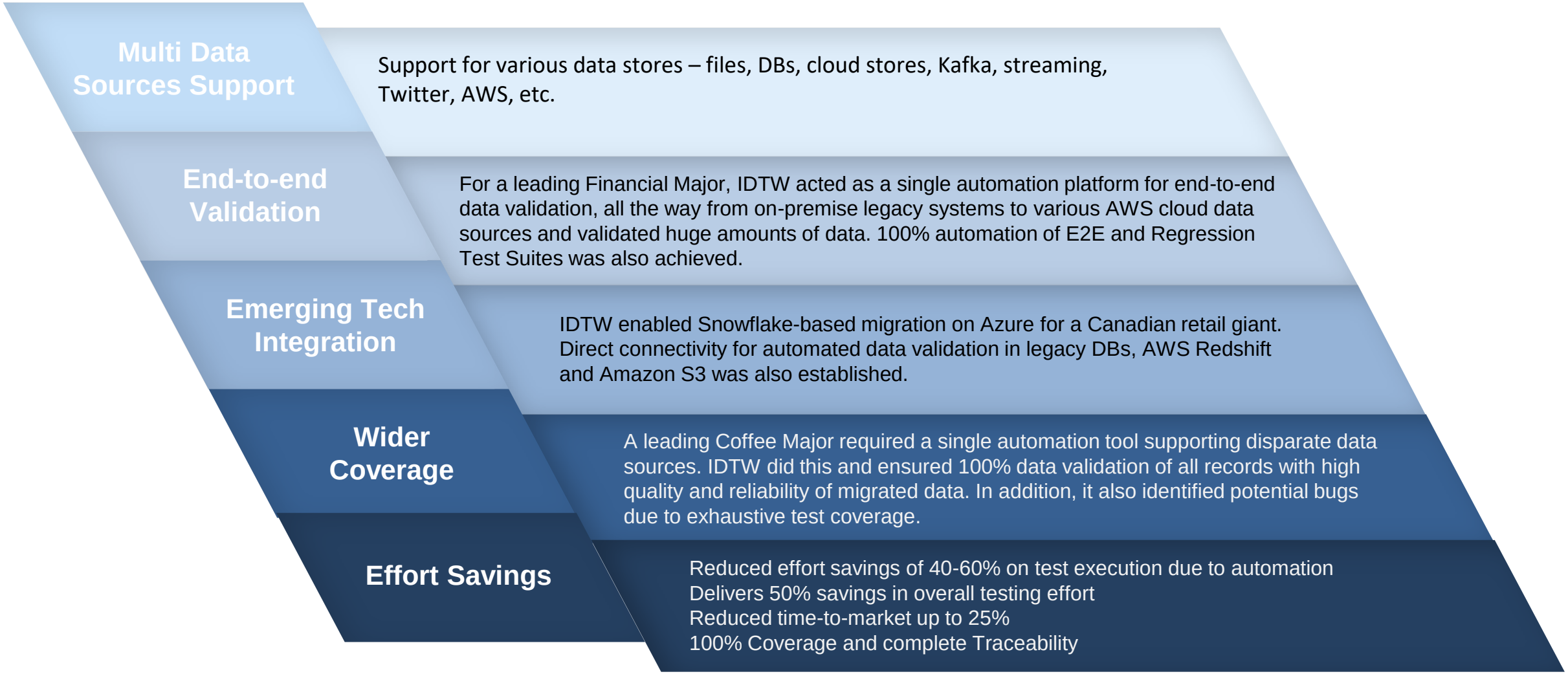
Statistical
- Column Statistics
- Outliers Check
- Data Duplication Check

Relationship
- Foreign Key Verification
- Cardinality Verification

Pattern
- Standardization Check
- Precision Constraints
- Format Verification

Custom Rules
- Domain-specific Business Rules

Benefits


Multi Data Sources Support
Support for various data stores – files, DBs, cloud stores, Kafka, streaming, Twitter, AWS, etc.
End-to-end Validation
For a leading Financial Major, IDTW acted as a single automation platform for end-to-end data validation, all the way from on-premise legacy systems to various AWS cloud data sources and validated huge amounts of data. 100% automation of E2E and Regression Test Suites was also achieved.
Emerging Tech Integration
IDTW enabled Snowflake-based migration on Azure for a Canadian retail giant. Direct connectivity for automated data validation in legacy DBs, AWS Redshift and Amazon S3 was also established.
Wider Coverage
A leading Coffee Major required a single automation tool supporting disparate data sources. IDTW did this and ensured 100% data validation of all records with high quality and reliability of migrated data. In addition, it also identified potential bugs due to exhaustive test coverage.
Effort Savings
Reduced effort savings of 40-60% on test execution due to automation
Delivers 50% savings in overall testing effort
Reduced time-to-market up to 25%
100% Coverage and complete Traceability

THANK YOU

© 2019 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.