DATA-CENTRIC QA ON CUSTOMER 360

PROJECT OVERVIEW

The retail giant manages a deluge of customer data from stores, POS devices, online transactions, items bought, and more. This amount of data was kept in various databases for at least five years and was organised by brands, transaction type (online or in-person), account management-related information, and operational data for specific products. Although data is operating in a "logical" silo, the retailer was unable to see an integrated customer view because it was difficult to map the appropriate data points from each database to feed into a customer intelligence engine.

SOLUTION DELIVERED

ETL testing and Big Data testing

KEY HIGHLIGHTS

- End to end ETL Process Data Validations
- Using a custom automation framework, the effort reduced from 8 hours to less than an hour

CLIENT DOMAIN

Retail, E-commerce

ABOUT CLIENT

Our client is a global Retail business with over $4 billion in revenue. The chain operates a warehouse of brands that are sold across extensive channels and in flexible models. The end customers can purchase tailored apparel from the available suite of brands, rent outfits or order custom apparel with individual specifications.

BUSINESS CHALLENGES

Given a Customer 360 perspective using the data, the retail company would analyze customer's journey and sales values for a dice of categories, transaction dynamics over the recent months, predict customer behaviors, forecast product demands and design right marketing campaigns. The Retail giant manages a flood of customer information at stores, POS units, online transactions, products purchased and so on. This volume of data, for at least over 5 years, was stored in different databases categorized by brands, mode of transactions (Digital or Store), account management related, product-wise operational data. Even though data is operating in a 'logical' silo, the Retailer was unable to visualize an integral customer view; mapping the right data points from each of the databases to feed into a Customer Intelligence engine was a challenge.

- Today, several Big Data ETL tools are black-box with respect to a GUI.
- This makes it challenging to manually assess the data quality after data transformations.
- This problem is further complicated when the data pipelines require data to be transformed in real time.
- This created a need to perform automated validations of such data pipelines to ensure data integrity

SOLUTION HIGHLIGHTS

- Indium Software's Team employed Retail subject matter / ETL expert to comprehend the data workflow, transformations, business objectives and the performance factors of the system.
- The team employed is Full-Stack for the requirement: Knowledge of Retail, ETL Expert, Big Data Technologies, Experience working in different Testing Environments, Test Tools proficiency.
- Our testers work with Data Scientists and Warehouse Architects for various projects, from which high-power domain understanding is carried down.
- Defined Automation scope for Reconciliation tests – each metric in the processing layer would require 15-30 queries to attempt testing; the overall testing scope required to cover 50 metrics which consumed 8 hours of manual effort. Using a custom automation framework, the effort reduced to less than an hour.
- Detailed Test Scenarios in standard templates applicable to each testing type.

TECH STACK

AGILONE   ORACLE

MySQL   UNIX

Data stage (ETL),
Defect Management Tools,

ETL Complexity

Data Sources	MDM, Customer data by Source Channel, Product DB
Data Source Types	Oracle DB, MySQL, Flat Files
Customer 360	Dynamic features for Data Target Holding
Transformation	80 Tables with simple-to-complex transformational nature

Test Strategy

Data sets were gathered in batches instead of mass volumes for Test Sprints. Estimations were done accordingly, and test results are produced for each batch.

Batch 0	1 month sample data (recent)
Batch 1	Historic data (5 years)
Batch 2	5 month data (recent)
Batch 3	1 month (most recent)
Batch 4	Incremental data (daily)

TEST COVERAGE MATRIX

Customer 360 ETL Transformation	Scope Of Testing

DATA AGGREGATION
Collective data is maintained in Customer360 using ETL Tools (stored in .CSV files) for better performance

- Production Validation Testing on data being moved to Customer360 for correct order of data
- Flow Tests

CUSTOMER IDENTITY
Data is cleansed and fed into required standard 'formats' before loading into Customer360 using match and merge rules, Identity Resolution Engine, De Duplication for customer information

QA Validations on the data for defined Formats and Layouts
- Source to Target Testing Data
- Acquisition Testing Compatibility

DATA INTEGRITY
Match back from Customer 360 to original Databases

- Data Completeness Tests for
- Transformation

DATA PROCESSING
Apply layered rules to derive/set the context for BI data

REAL TIME DATA
Data flux is real time for daily/weekly campaigns

Scope Of Testing

Validate the processed data in a temporary database against the rules and context of customer intelligence Data Quality Validations (syntax and reference tests)

- Metadata Testing
- IRE Validations/ De duplication
- Reconciliation Testing and
- Automation
- UAT

- Incremental ETL Testing
- Verify data updates from the sources to the aggregate Customer360

ABOUT INDIUM

Indium is a Digital Engineering Services leader and Full Spectrum Integrator that helps customers embrace and navigate the Cloud-native world with Certainty. With deep expertise across Applications, Data & Analytics, AI, DevOps, Security and Digital Assurance we "Make technology work" and accelerate business value, while adding scale and velocity to customer's digital journey on AWS.

indium
Make Technology Work

USA	INDIA	UK	SINGAPORE
Cupertino | Princeton	Chennai | Bengaluru | Mumbai | Hyderabad	London	Singapore
Toll-free: +1-888-207-5969	Toll-free: 1800-123-1191	Ph: +44 1420 300014	Ph: +65 6812 7888

www.indiumsoftware.com

For Sales Inquiries
sales@indiumsoftware.com

For General Inquiries
info@indiumsoftware.com