Application:
Customer 360 and ETL

Services Offered:
ETL Testing and Big Data QA

Tools:
Data stage (ETL), AgilOne (ETL), Oracle and MySQL DB, Defect Management Tools, UNIX

Key Highlights
End to end ETL Process Data Validations

Domain:
Retail

Team:
1 test lead and 1 Sr.ETL test expert

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.
### Status Quo

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.

### Business Problem

The business case was resolved by implementing a Customer 360 interface. Customer’s data in terms of transaction value, mode of shopping, recent shopping patterns and so on, are captured into Customer 360 using ETL processes. However, the integrity of the data from sources to the ETL drill and into the customer 360 database was unsubstantiated. Hence the need for a dedicated ETL Testing process.

### Solution Approach

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.

![Illustration of ETL process and content aggregation in Customer 360](image)

#### ETL Complexity

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>MDM, Customer data by Source Channel, Product DB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Source Types</td>
<td>Oracle DB, MySQL, Flat Files</td>
</tr>
<tr>
<td>Customer 360</td>
<td>dynamic features for Data Target Holding</td>
</tr>
<tr>
<td>Transformation</td>
<td>80 Tables with simple-to-complex transformational nature</td>
</tr>
</tbody>
</table>
Test Coverage Matrix

**Customer 360 ETL Transformation**

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

**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.

**DATA INTEGRITY**
Match back from Customer 360 to original Databases.

**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**
- Production Validation Testing on data being moved to Customer360 for correct order of data
- Flow Tests
- QA Validations on the data for defined Formats and Layouts
  - Source to Target Testing
  - Data Acquisition Testing
  - Compatibility
- Data Completeness Tests for Transformation
- 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

**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)
5 Value Delivered

» 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