# SOURCE-TO-TARGET MAPPING DOCUMENT

E-commerce Data Migration Project

**PROJECT INFORMATION**

| **Field** | **Details** |
| --- | --- |
| Project Name: | E-commerce Platform Migration |
| Version: | 1.0 |
| Date: | December 2024 |
| Author: | Data Analytics Team |
| Source System: | Legacy E-commerce DB (MySQL) |
| Target System: | Modern Data Warehouse (PostgreSQL) |
| Migration Type: | ETL Pipeline |
| Data Volume: | ~2.5M Customer Records |

**1. CUSTOMER DATA MAPPING**

**Business Rules for Customer Data:**

* All customer emails must be unique and validated
* Phone numbers should be standardized to international format
* Customer status should default to 'Active' for new records
* Registration date cannot be future dated
* Customer tier is calculated based on total purchase amount

**Customer Mapping Table:**

| **Source**  **Field** | **Source**  **Table** | **Source**  **Data Type** | **Target**  **Field** | **Target**  **Table** | **Target**  **Data Type** | **Transformation**  **Logic** | **Validation**  **Rules** | **Mandatory** | **Status** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cust\_id | customers | INT(11) | customer\_id | dim\_customer | BIGINT | Direct mapping | Primary Key, Auto-increment | Yes | Active |
| first\_name | customers | VARCHAR(50) | first\_name | dim\_customer | VARCHAR(100) | TRIM() + INITCAP() | Not null, Length > 1 | Yes | Active |
| last\_name | customers | VARCHAR(50) | last\_name | dim\_customer | VARCHAR(100) | TRIM() + INITCAP() | Not null, Length > 1 | Yes | Active |
| email\_address | customers | VARCHAR(100) | email | dim\_customer | VARCHAR(255) | LOWER() + Email validation | Unique, Valid email format | Yes | Active |
| phone\_num | customers | VARCHAR(20) | phone\_number | dim\_customer | VARCHAR(25) | Format to +1-XXX-XXX-XXXX | Valid phone format | No | Active |
| birth\_date | customers | DATE | date\_of\_birth | dim\_customer | DATE | Direct mapping | Age >= 13, Date <= TODAY | No | Active |
| registration\_dt | customers | DATETIME | registration\_date | dim\_customer | TIMESTAMP | Convert to UTC timezone | Not null, <= CURRENT\_DATE | Yes | Active |
| status | customers | CHAR(1) | customer\_status | dim\_customer | VARCHAR(20) | A='Active', I='Inactive', S='Suspended' | Must be valid status code | Yes | Active |

**2. PRODUCT DATA MAPPING**

**Business Rules for Product Data:**

* Product SKU must be unique across all products
* Product prices must be greater than 0
* Category hierarchy should be maintained (Category > Subcategory)
* Discontinued products should retain historical data
* Product descriptions should be sanitized for HTML/special characters

**Product Mapping Table:**

| **Source Field** | **Source Table** | **Source Data Type** | **Target Field** | **Target Table** | **Target Data Type** | **Transformation Logic** | **Validation Rules** | **Mandatory** | **Complexity** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **product\_id** | **products** | **INT(11)** | **product\_id** | **dim\_product** | **BIGINT** | **Direct mapping** | **Primary Key, Auto-increment** | **Yes** | **Low** |
| **sku** | **products** | **VARCHAR(50)** | **product\_sku** | **dim\_product** | **VARCHAR(100)** | **UPPER() + Remove spaces** | **Unique, Not null** | **Yes** | **Low** |
| **product\_name** | **products** | **VARCHAR(200)** | **product\_name** | **dim\_product** | **VARCHAR(255)** | **TRIM() + HTML decode** | **Not null, Length > 3** | **Yes** | **Medium** |
| **description** | **products** | **TEXT** | **product\_description** | **dim\_product** | **TEXT** | **Sanitize HTML + Character encoding** | **Safe content only** | **No** | **High** |
| **price** | **products** | **DECIMAL(10,2)** | **unit\_price** | **dim\_product** | **DECIMAL(12,2)** | **Round to 2 decimal places** | **Must be > 0** | **Yes** | **Low** |
| **category\_id** | **products** | **INT(11)** | **category\_id** | **dim\_product** | **BIGINT** | **Lookup category hierarchy** | **Must exist in category table** | **Yes** | **Medium** |
| **inventory\_count** | **products** | **INT(11)** | **stock\_quantity** | **dim\_product** | **INTEGER** | **Direct mapping** | **Must be >= 0** | **Yes** | **Low** |
| **is\_active** | **products** | **TINYINT(1)** | **product\_status** | **dim\_product** | **VARCHAR(20)** | **1='Active', 0='Discontinued'** | **Must be valid status** | **Yes** | **Low** |

**3. ORDER TRANSACTION MAPPING**

**Business Rules for Order Data:**

* Order total must equal sum of line items plus tax and shipping
* Order date cannot be in the future
* Customer must exist before order can be created
* Order status follows specific workflow: Pending → Processing → Shipped → Delivered
* Payment status must be validated against payment gateway

**Order Mapping Table:**

| **Source Field** | **Source Table** | **Source Data Type** | **Target Field** | **Target Table** | **Target Data Type** | **Transformation Logic** | **Validation Rules** | **Mandatory** | **Status** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| order\_id | orders | INT(11) | order\_id | fact\_sales | BIGINT | Direct mapping | Primary Key | Yes | Active |
| customer\_id | orders | INT(11) | customer\_key | fact\_sales | BIGINT | Foreign key lookup | Must exist in dim\_customer | Yes | Active |
| order\_date | orders | DATETIME | order\_date | fact\_sales | DATE | Extract date part only | Not null, <= CURRENT\_DATE | Yes | Active |
| total\_amount | orders | DECIMAL(10,2) | order\_total | fact\_sales | DECIMAL(12,2) | Round to 2 decimals | Must be > 0 | Yes | Active |
| tax\_amount | orders | DECIMAL(8,2) | tax\_amount | fact\_sales | DECIMAL(10,2) | Direct mapping | Must be >= 0 | No | Active |
| shipping\_cost | orders | DECIMAL(8,2) | shipping\_amount | fact\_sales | DECIMAL(10,2) | Direct mapping | Must be >= 0 | No | Active |
| order\_status | orders | VARCHAR(20) | order\_status | fact\_sales | VARCHAR(30) | Standardize status codes | Valid status values only | Yes | Active |

**4. DATA QUALITY METRICS**

| **Table** | **Total Records** | **Valid Records** | **Data Quality %** | **Common Issues** | **Remediation Action** |
| --- | --- | --- | --- | --- | --- |
| customers | 2,456,789 | 2,398,124 | 97.6% | Invalid email formats, Missing phone numbers | Email validation, Phone standardization |
| products | 156,234 | 155,891 | 99.8% | HTML in descriptions, Negative prices | HTML sanitization, Price validation |
| orders | 8,967,234 | 8,923,456 | 99.5% | Future dates, Orphaned customer IDs | Date validation, Referential integrity checks |

**IMPLEMENTATION NOTES**

**Migration Schedule:** The data migration will be executed in phases over 3 weekends to minimize business impact.

**Rollback Plan:** Complete database backup will be taken before migration with automated rollback procedures.

**Validation Process:** Post-migration data validation will include row counts, data sampling, and business rule verification.

**Performance Considerations:** Bulk insert operations will be used with batch sizes of 10,000 records to optimize performance.

**SIGN-OFF AND APPROVALS**

**Data Architecture Team:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_

**Business Stakeholder:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_

**IT Operations:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_

**Quality Assurance:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_