

Product Specification Document: Shipment Manifest Data Processing System

Objective

To build a structured, scalable, and continuously improvable data pipeline for cleaning, enriching, and consolidating shipment manifest files into a central table, enabling better visibility, analytics, and downstream integrations.

Supported File Input

- Format: CSV
- Typical Columns: Master BOL, House BOL, Container Numbers, Commodity, Carrier Code, Package Type, Shipper, Consignee, Notify Party 1, Notify Party 2, Shipper Address, Consignee Address, Place of Receipt, Foreign Port of Lading, US Port of Unlading, Gross Weight, Net Weight, Number of Packages, etc.

Processing Steps

1. Remove Exact Duplicates: Eliminate rows that are an exact match across all columns.
2. De-duplicate by MBL + Container Number:
 - All rows have HBL retain and mark LCL = Yes
 - If only one row lacks HBL among two remove the one without
 - More than two rows keep those with HBL, mark LCL = Yes
3. Map SCAC to LSP: Use SCAC mapping to derive full LSP name.
4. Standardize Party Names: Normalize suffixes like LTD. LIMITED, PVT. PRIVATE.
5. Standardize Place of Receipt: Normalize port names like NHAVA SHEVA variants.
6. Extract Shipper City: Match from address using Indian city list.
7. Extract Consignee City: Match from address using US/Canada city list.
8. Entity Classification: Classify entities into 8 categories (e.g., NVOCC, Exporter, Importer).
9. Unified LSP: Use LSP if HBL exists; otherwise fallback to matching entity roles.
10. Classify Commodity: Rule-based (e.g., AUTO PARTS Automotive Parts). ML extension possible.
11. Extract HS Code: Regex extract from Commodity text, normalize format.
12. Normalize Units of Measurement (Planned): Extract PKGS, CTNS, PCS, etc.
13. Detect Dangerous Goods (Planned): Flag BATTERY, FLAMMABLE, etc.

Product Specification Document: Shipment Manifest Data Processing System

14. Standardize Ports (Planned): Clean Foreign Port of Lading & US Port of Unlading.

Central Table Logic

Each cleaned file is appended to a central table. If a row is already present (exact match), it is skipped.

Output Columns (Sample)

- LCL: Flag for Less-than-Container Load
- LSP: Mapped name from SCAC code
- Shipper City / Consignee City: Extracted from address
- Shipper/Consignee/LSP Category: Classified entity type
- Unified LSP: Final logistics partner after fallback logic
- Commodity Category: Enriched product group label
- Extracted HS Code: Parsed HS code from commodity text

Evolution Plan

- Future enhancements include:
- ML-based classification for Commodity and Entity
 - Tariff/HS Code validation
 - Reporting layer and dashboard integration