**Transit Data Warehouse Implementation Documentation**

**IST 722: Data Warehouse Project**

**Team: The Special Ones (Team 1)**

## 1. Sample Source to Target Mapping

### Staging Layer

|  |  |  |
| --- | --- | --- |
| Source | Target | Transformations |
| CSV.Fact\_ID | RAW\_TRANSIT\_DATA.Fact\_ID | Direct Load |
| CSV.Datetime | RAW\_TRANSIT\_DATA.Datetime | Convert to TIMESTAMP |
| CSV.Stop\_ID | RAW\_TRANSIT\_DATA.Stop\_ID | Direct Load |

### Dimension Tables

|  |  |  |
| --- | --- | --- |
| Source | Target | Transformations |
| stg\_raw\_transit.datetime,hour,day,month,year | DIM\_TIME | Generate time\_key, calculate is\_weekend |
| stg\_raw\_transit.stop\_\* | DIM\_STOPS | Direct mapping with deduplication |
| stg\_raw\_transit.line\_\* | DIM\_LINES | Direct mapping with deduplication |
| stg\_raw\_transit.remote\_unit\_\* | DIM\_REMOTE\_UNITS | Direct mapping with deduplication |

### Fact Tables

|  |  |  |
| --- | --- | --- |
| Source | Target | Transformations |
| stg\_raw\_transit | FACT\_RIDER\_VOLUME | Join with dimensions, calculate utilization |
| stg\_raw\_transit | FACT\_STOP\_UTILIZATION | Aggregate by stop, calculate metrics |
| stg\_raw\_transit | FACT\_LINE\_PERFORMANCE | Calculate performance metrics |
| stg\_raw\_transit | FACT\_REMOTE\_MONITOR | Track events and status |
| stg\_raw\_transit | FACT\_RIDER\_SEGMENTS | Segment analysis and categorization |

## 2. Loading Sequence

**1. Initial Load:**

- Load CSV to RAW\_TRANSIT\_DATA

- Transform to staging view

- Load dimension tables

- Load fact tables

**2. Incremental Updates:**

- Append new data to RAW\_TRANSIT\_DATA

- Update staging view

- Update dimension tables (SCD Type 1)

- Append to fact tables

## 3. Data Quality Checks

**1. Staging Layer:**

- No null keys

- Valid dates

- Valid numeric values

**2. Dimension Tables:**

- Unique keys

- No orphaned references

- Complete attributes

**3. Fact Tables:**

- Valid foreign keys

- Business rule validations

- Metric range checks

## 4. Implementation Standards

**1. Naming Conventions:**

- Staging: stg\_\*

- Dimensions: dim\_\*

- Facts: fact\_\*

- Keys: \*\_id, \*\_key

**2. Data Types:**

- Keys: INTEGER

- Dates: TIMESTAMP

- Text: VARCHAR

- Metrics: FLOAT/INTEGER

**3. Architecture:**

- Three-layer architecture (STAGING, ODS, DWH)

- Star schema design

- Conformed dimensions