

Data Extraction and Load Procedures Document Sample

1. Overview

This document outlines the key procedures and steps for extracting and loading data between source and target systems. The intent is to ensure accurate, efficient, and repeatable data movement processes.

2. Source and Target Details

Source System	ERP Database (MySQL)
Target System	Data Warehouse (PostgreSQL)
Extraction Frequency	Daily at 01:00 AM UTC
Load Window	01:30 AM - 02:00 AM UTC

3. Extraction Procedure

1. Connect to the source ERP database using secure credentials.
2. Execute the following query to retrieve necessary records:

```
SELECT id, name, amount, updated_at
FROM sales_orders
WHERE updated_at >= CURDATE();
```

3. Export query results to a CSV file, named as *sales_orders_YYYYMMDD.csv*.
4. Validate record count and ensure file integrity.
5. Store the file in the designated extraction directory: */data/extract/*.

4. Load Procedure

1. Transfer the extracted file to the target server via SFTP.
2. Connect to the data warehouse using secure credentials.
3. Stage data into a temporary table: *stg_sales_orders*.
4. Perform transformation/mapping if required (e.g., data type conversions, field mapping).
5. Validate data integrity and referential constraints.
6. Insert records from staging table into the target table: *dw_sales_orders*.
7. Archive the processed file in */data/archive/*.

5. Error Handling & Logging

- Log all extraction and load activities in *etl_process.log*.
- Notify data engineering team if errors are detected during extraction or loading.
- Retries up to 3 times if load fails due to connectivity issues.

6. Contact

Data Engineering Team
Email: data-team@example.com