

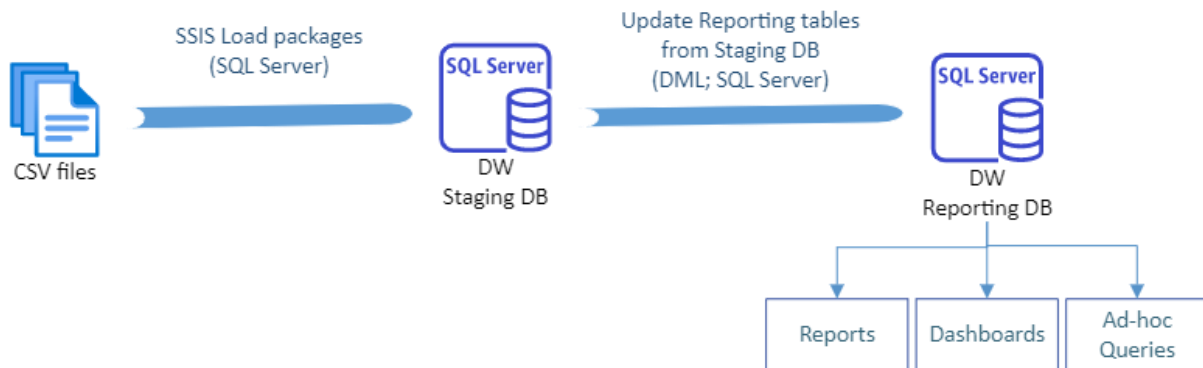
Design for importing data into a data warehouse

This document contains design principles for implementation of the vehicle's usage data ETL process into the data warehouse.

1. Basic Architecture

Tools chosen for the implementation:

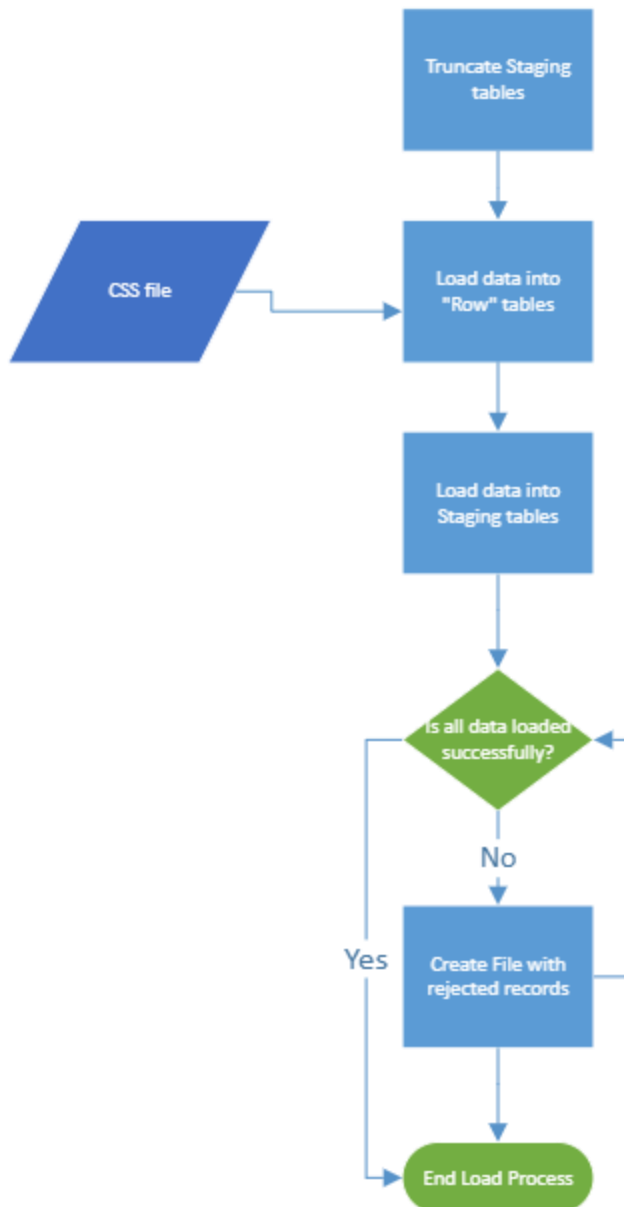
- SQL Server for DW database
- SSIS for data loading from CSS files into the staging database
- DML for populating reporting database from staging



2. Pipelines

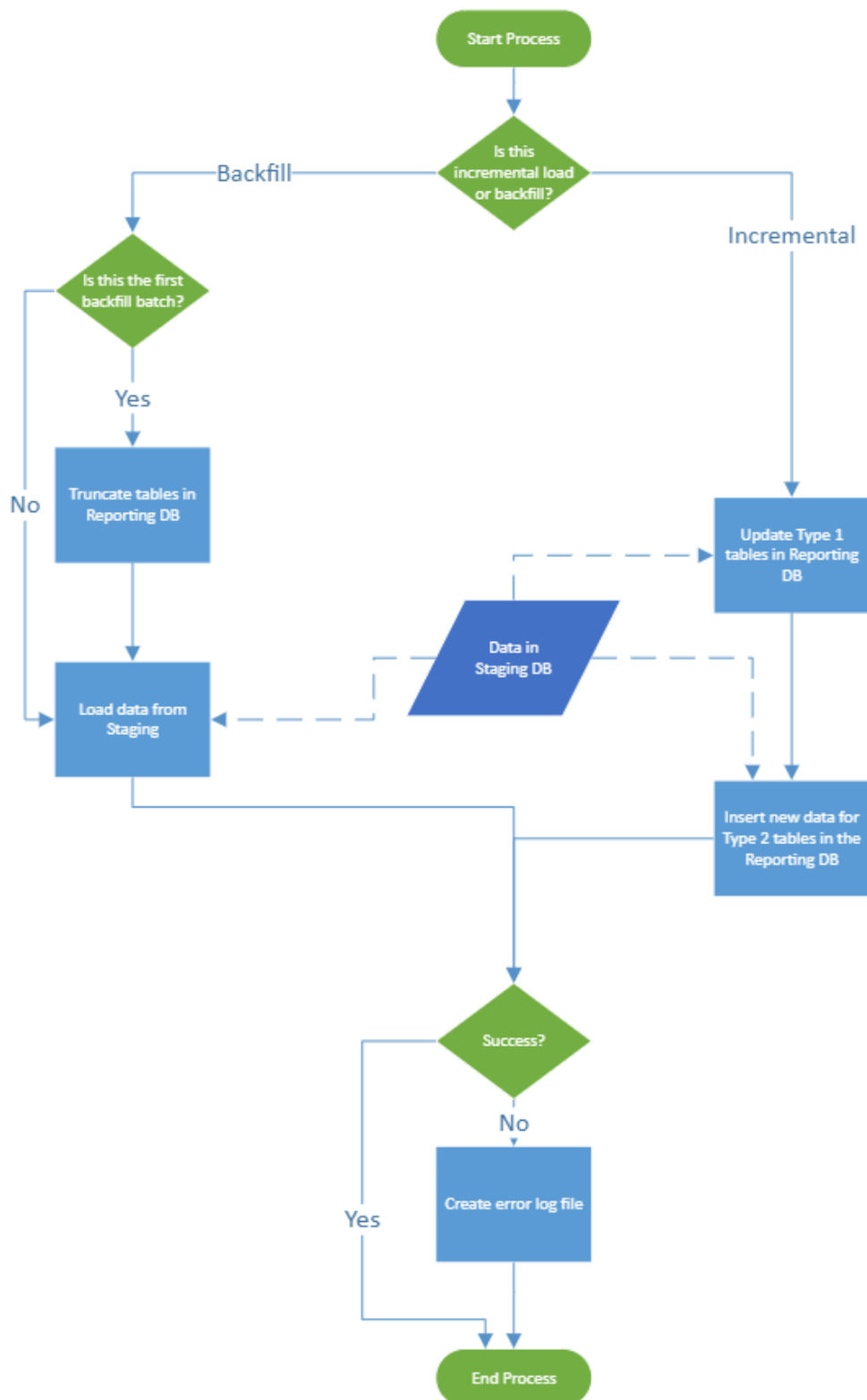
Below are the flowcharts for 2 processes: initial loading into the Staging database, and data load from Staging to Reporting.

Load Into Staging database

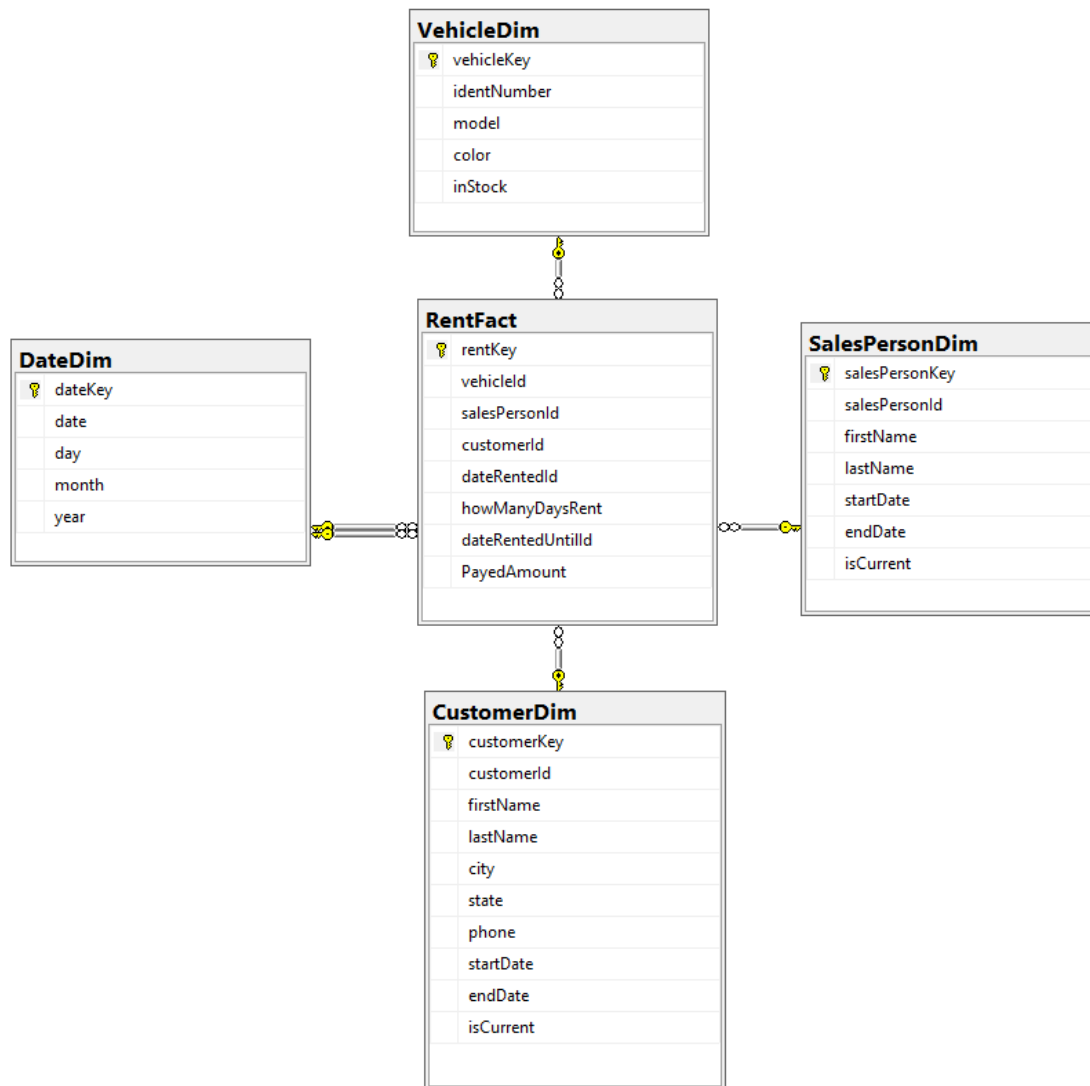


"Row" tables are historical tables that contain row data from CSS files (they are optional).

Load Into Reporting database



3. Example of Data Model



4. Example of Data Load and Transformation

This is an example of data load and transformation from Customer CSV file into the Staging database, and into Reporting database.

CSV File: Customer
1003, Bob, Martin, Austin, TX, 512-487-7777

Staging DB

Customer	
id	1003
firstName	Bob
lastName	Martin
City	Austin
State	TX
Phone	512-487-7777

Customer_IdLookup	
sourceType	CSV
source	Customer
id	1003
dwKey	205

Reporting DB before ETL

Customer	
customerKey	20005
customerId	205
firstName	Bob
lastName	Martin
city	Austin
state	TX
phone	5122582457
startDate	5/1/2010
endDate	12/31/2099
isCurrent	1

Reporting DB after ETL

Customer		
customerKey	20005	38025
customerId	205	205
firstName	Bob	Bob
lastName	Martin	Martin
city	Austin	Austin
state	TX	TX
phone	5122582457	5124877777
startDate	5/1/2010	2/10/2021
endDate	2/9/2021	12/31/2099
isCurrent	0	1

5. Performance and scalability considerations

- Consider a mix of rowstore and columnstore indexes for performance benefit
- Partitioning big tables and views for performance
- Use fast storage
- Replication, High Availability groups, clustering in SQL Server
- Use linked servers for executing distributed queries
- Using middleware to route queries to correct database if DW is broken down into several databases (ex. US location, Europe etc.)

6. Security considerations

- Login with read/write permissions on the tables in Staging and Reporting DBs for ETL process

- Login with read/delete files permissions in the file system where CSV files, archived CSV file and log files are located
- Encryption of sensitive information in the DB
- Roles for BI users with read permissions in Reporting DB on only data that they need

7. Monitoring and alerting

- Log files, and files with rejected records created during ETL
- Email on errors during ETL
- GUI application where ETL status, errors and warnings can be displayed