DSE 203 Final
California Highway Traffic Analysis

Conway Wong
John Gill III
Kevin Dyer

December 12, 2015

Purpose

Combine Traffic, Weather, and CHP Incident data into a single data warehouse, enabling systems to determine traffic rates and CHP incidents with respect to weather; and potentially other determinations

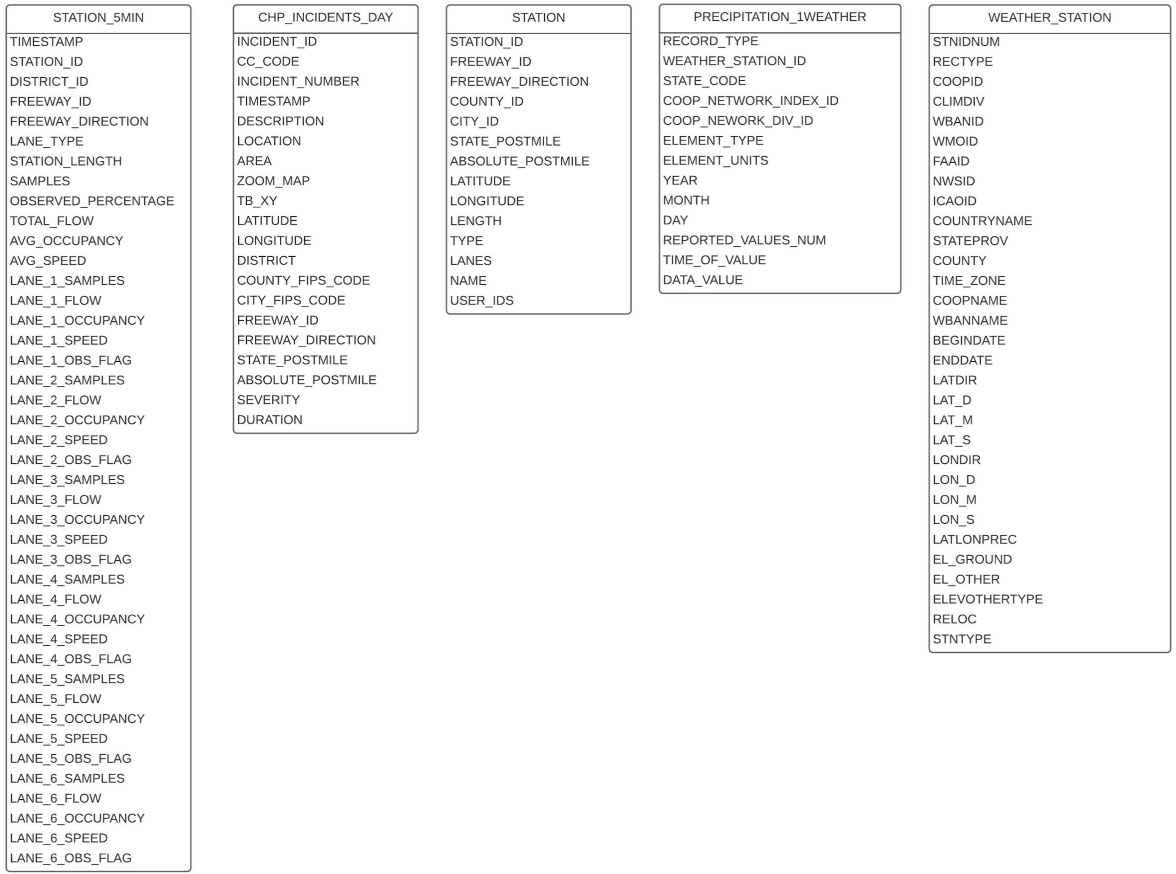
Data Sources

- Caltrans Performance Measurement System (PeMS)
- National Oceanic and Atmospheric Administration
- California Highway Patrol

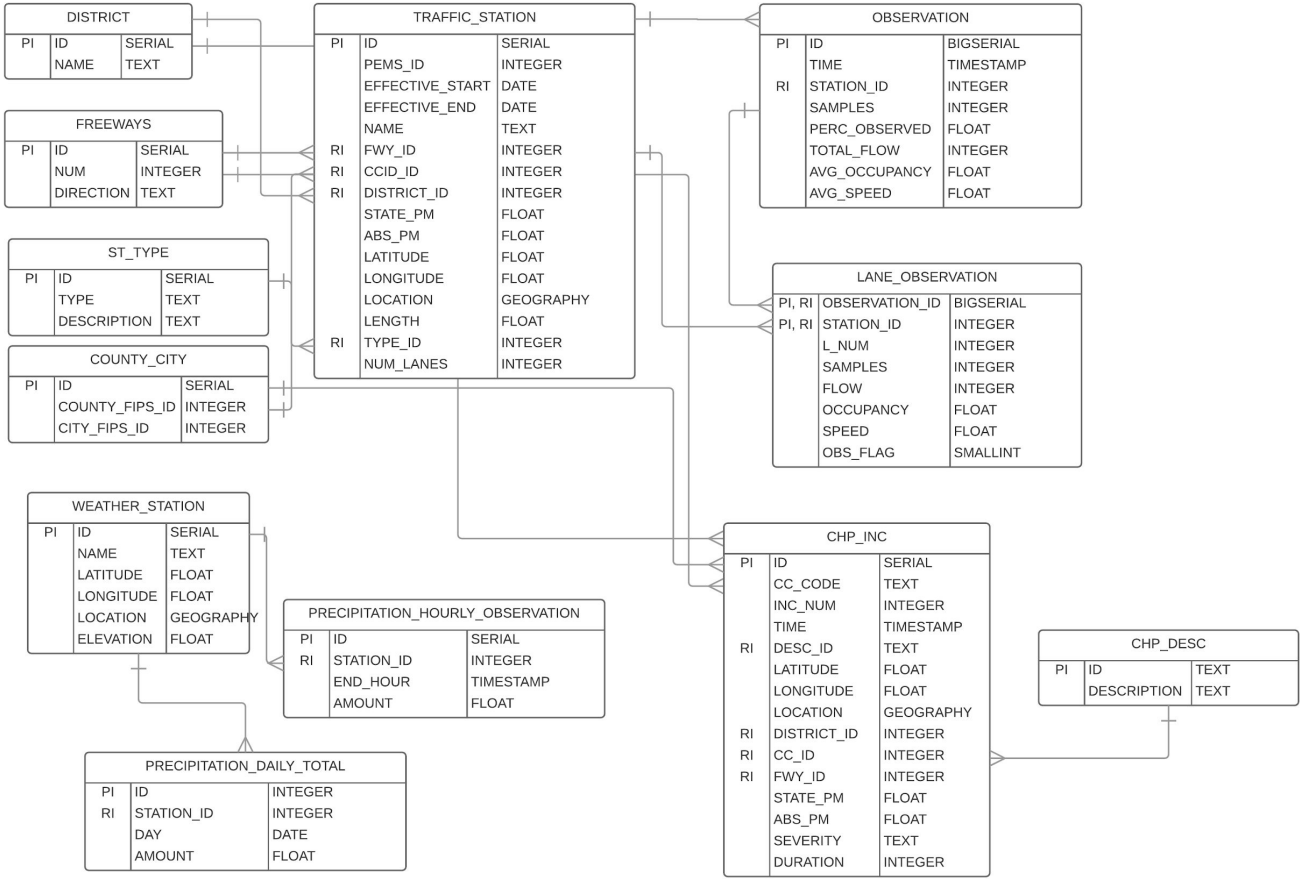
Data Integration Tasks

- Creation of Data Warehouse Schema
- Obtain Traffic Data from PEMs
- Obtain Weather Station Data from NOAA
- Obtain Weather Precipitation Data from NOAA
- Obtain California Highway Patrol Incident Data from CHP
- Execute Queries against Data Warehouse Schema

Source Schema



Target Schema



Clover ETL Graph Demo

Queries

- What traffic station has the largest difference in average speed over the first two weeks of the month?
- How significant is the difference in traffic throughput on a rainy Monday vs a non-rainy Monday?
- Does trace amount of precipitation affect the number of CHP traffic incidents on a given day? (Trace precipitation is defined as a weather station registering precipitation but less than the unit granularity of the sensor)
- Identify the top 5 freeways with respect to traffic speed.
- Is the traffic throughput of one freeway indicative of others?

IPython Notebook Query Results Demo

Visualization