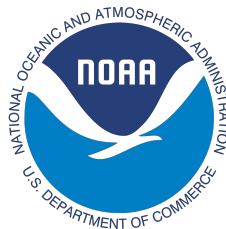


DSE 203 Final

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

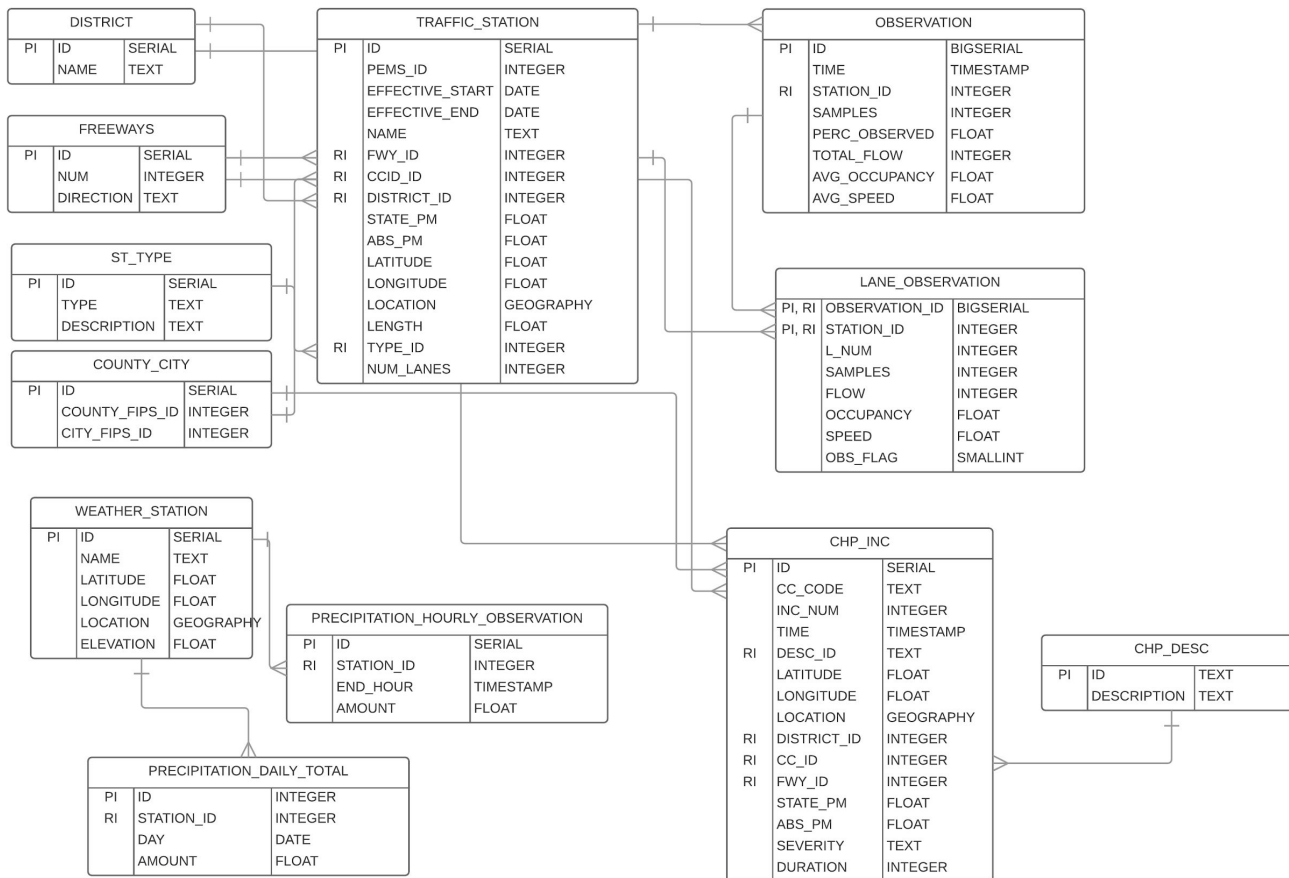
Data Integration Tasks

- 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

STATION_5MIN	CHP_INCIDENTS_DAY	STATION	PRECIPITATION_1WEATHER	WEATHER_STATION
TIMESTAMP	INCIDENT_ID	STATION_ID	RECORD_TYPE	STNIDNUM
STATION_ID	CC_CODE	FREEWAY_ID	WEATHER_STATION_ID	RECTYPE
DISTRICT_ID	INCIDENT_NUMBER	FREEWAY_DIRECTION	STATE_CODE	COOPID
FREEWAY_ID	TIMESTAMP	COUNTY_ID	COOP_NETWORK_INDEX_ID	CLIMDIV
FREEWAY_DIRECTION	DESCRIPTION	CITY_ID	COOP_NETWORK_DIV_ID	WBANID
LANE_TYPE	LOCATION	STATE_POSTMILE	ELEMENT_TYPE	WMOID
STATION_LENGTH	AREA	ABSOLUTE_POSTMILE	ELEMENT_UNITS	FAAID
SAMPLES	ZOOM_MAP	LATITUDE	YEAR	NWSID
OBSERVED_PERCENTAGE	TB_XY	LONGITUDE	MONTH	ICAID
TOTAL_FLOW	LATITUDE	LENGTH	DAY	COUNTRYNAME
AVG_OCCUPANCY	LONGITUDE	TYPE	REPORTED_VALUES_NUM	STATEPROV
AVG_SPEED	DISTRICT	LANES	TIME_OF_VALUE	COUNTY
LANE_1_SAMPLES	COUNTY_FIPS_CODE	NAME	DATA_VALUE	TIME_ZONE
LANE_1_FLOW	CITY_FIPS_CODE	USER_IDS		COOPNAME
LANE_1_OCCUPANCY	FREEWAY_ID			WBANNAME
LANE_1_SPEED	FREEWAY_DIRECTION			BEGINDATE
LANE_1_OBS_FLAG	STATE_POSTMILE			ENDDATE
LANE_2_SAMPLES	ABSOLUTE_POSTMILE			LATDIR
LANE_2_FLOW	SEVERITY			LAT_D
LANE_2_OCCUPANCY	DURATION			LAT_M
LANE_2_SPEED				LAT_S
LANE_2_OBS_FLAG				LONDIR
LANE_3_SAMPLES				LON_D
LANE_3_FLOW				LON_M
LANE_3_OCCUPANCY				LON_S
LANE_3_SPEED				LATLONPREC
LANE_3_OBS_FLAG				EL_GROUND
LANE_4_SAMPLES				EL_OTHER
LANE_4_FLOW				ELEVOTHERTYPE
LANE_4_OCCUPANCY				RELOC
LANE_4_SPEED				STNTYPE
LANE_4_OBS_FLAG				
LANE_5_SAMPLES				
LANE_5_FLOW				
LANE_5_OCCUPANCY				
LANE_5_SPEED				
LANE_5_OBS_FLAG				
LANE_6_SAMPLES				
LANE_6_FLOW				
LANE_6_OCCUPANCY				
LANE_6_SPEED				
LANE_6_OBS_FLAG				

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