Real-time Weather Monitor Database

Database Systems CS4347.004

Daniel Crawford
Austin Bowden
Andrew Koenig
Lex Darocy

Problem Domain and Solution

Our group would like to build a database that reflects real-time measured weather conditions gathered from various automated weather stations scattered throughout the DFW Metroplex. The data will be split into two primary categories: measured data and derived data. Measured data will consist of tangible data, such as temperature, humidity, wind vectors, and rain accumulation. The derived products will consist of data that can be interpolated or calculated from the measured products. Derived products consist of data such as dewpoint, rain rate, historical high and low data, average wind speed, and long-term rain accumulation.

Identify and describe major entity types

Major Entity Types:

- Weather
- Wind Factor
- Weather Sensor
- Location
- Weather Station
- Historical Data

The data tables will be created as follows:

1. The data archive shall include all measured data:
    a. Temperature
    b. Humidity
    c. Wind vector
    d. Rain accumulation (high-res)
    e. Pressure
    f. Solar radiation
2. The real-time derived products table shall be a SQL view that includes these derived products:
    a. Dewpoint
    b. Heat index
    c. Wind chill
    d. Rain rate
3. A daily historical table shall be a SQL view that establishes these derived products:
    a. High/low values of these columns:
        i. Temperature
        ii. Humidity
        iii. Pressure
        iv. Wind speed
        v. Solar radiation
        vi. Heat index
        vii. Rain rate (high only)

    b. Timestamps of the high/low occurrence of the columns listed above
    c. Daily rain accumulation
4. A weekly historical table shall be a SQL view that calculates similar derived products to the daily historical table, but over a week-long period
5. A monthly historical table shall be a SQL view that calculates similar derived products to the daily historical table, but over a month-long period
6. A yearly historical table shall be a SQL view that calculates similar derived products to the daily historical table, but over a year-long period

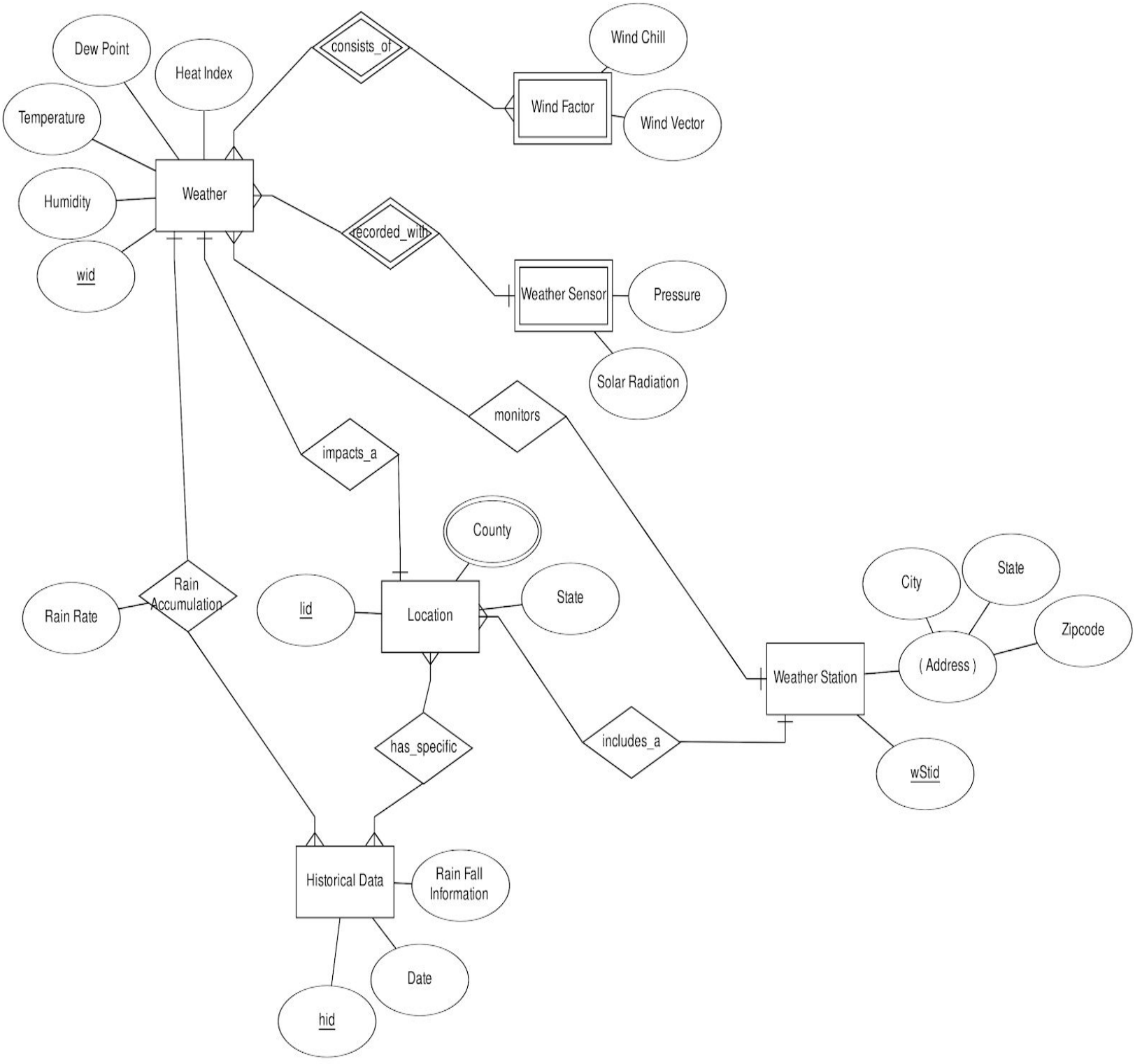
Assumptions about the business rules, and possible new entity types and relationships.
WIth the climate science evolving to consider additional sources of change, the tables in this project must be flexible to allow new measurements to be added. A good example of a new entity type is soil moisture levels. This is not a common measurement on weather stations, but the data can be useful for predicting and observing flooding patterns and agricultural water needs. The intrinsic structure of our database will allow these types of measurements to be added with minimal changes to the tables. The data flow of root measurements, then derived products, then reports can encompass most new measurements as new sensors are developed.

Assumptions on volumes of data
Weather data is collected at a given interval; this interval is typically referred to as the archive interval. Depending on the quality of the weather station and the expected rate of change, this interval will typically vary between 1 minute and 30 minutes. The faster reporting, at 1 minute intervals, can be useful for real-time analysis of weather conditions. The total number of records annually collected per weather station will exceed 500,000 records, but the data is simple and will not use a lot of disk space. The data types are all integers or floating point values, which can be stored very efficiently.

Data will be collected from weather stations in the DFW Metroplex. The data must be run through sanity checks to verify integrity; for example a temperature change greater than 5°/min would be unrealistic and allow us to see potential faulty data points. The derived products will be calculated internally within the DBMS using SQL views. This ensures that any data accessed will be up-to-date and not require a scheduled reporting task to reflect current conditions.

Summary
Gathering and collecting weather data and storing it into a database can be used to help better understand weather patterns and help better predict weather to be used by future applications. With enough data collected, it can used as historical data for advanced pattern recognition (APR) software to build models and provide early detections of weather changes. We will provide to the end-users a way to access weather data accumulated in the DFW metroplex and the ability to see different weather data around the complex. It can provide environmentalists with weather patterns and how weather fluctuates in the respected region. This can be used to help better understand weather patterns and changes within a specific region.

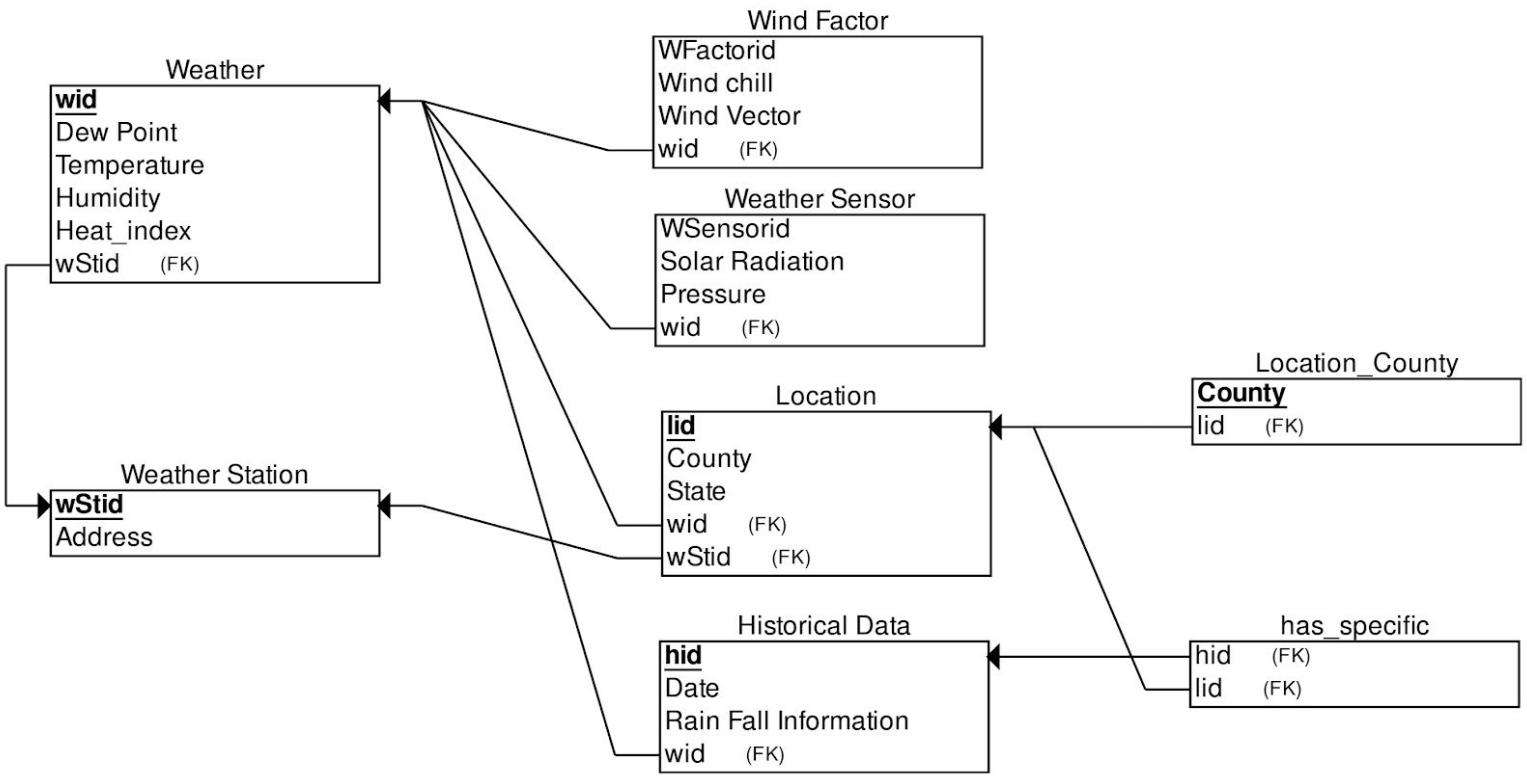
1. ERD Diagram



2. ERD Dictionary

Entity Type	Entity Type Description	Attributes	Keys	Relationships
Weather	A snapshot of weather conditions at a point in time	wid, humidity, temperature, dew point, heat index	wid	Weather consists of the base measurements (attributes) and the Wind Factor. Weather is recorded with a weather sensor. Weather impacts a location. In addition, weather is monitored with a weather station. A view of the weather is rain accumulation
Wind Factor	A snapshot of wind measurements	wind chill, wind vector		The wind factor is made of the wind vector and wind chill
Weather Sensor	A subset of sensors used to monitor weather conditions	pressure, solar radiation		Raw weather sensors record data
Weather Station	A weather station is the sum of instruments used to measure weather conditions	wStid, Address [City, State, Zipcode]	wStid	The weather station has an address consisting of a street address, city, state and zip code
Rain Accumulation	Instantaneous rain rate can be measured at a weather station	rain rate		
Location	A location is a general area	lid, county, state	lid	A location includes a weather station
Historical Data	Historical data is view of rain totals in recent history	hid, rainfall information, date	hid	A location has specific historical data, and weather has a historical rain accumulation

3. Schema Diagram



4. Schema Dictionary

Type	Table Description	Primary Key	SuperKey	Attributes			Foreign Keys
				Attribute	Data Type	Domain	
Weather	The weather in an area. Keeps track of all the details of what the current weather is.	wid	(wid, wStid, Dew Point, Temperature, Humidity, Heat_index)	wid	String	unique	
				Dew Point	INT	0 C to 40 C	
				Temperature	INT	Must be greater than -273, and is recorded in celsius	
				Humidity	INT	0% to 100%	
				Heat_index	INT	27 C to 58 C	
Weather Station	Where the data is being gathered from	wStid	(wStid, Address)	wStid	String	unique	
				Address	String	Must be a valid location	
Wind Factor	Details about the wind in an area		(WFactorid, wid, Wind Chill, Wind Vector)	Wfactorid	String	unique	wid
				Wind Chill	INT	Must be greater than -273, and is recorded in celsius	
				Wind Vector	INT	0 to 360 degrees of rotation	
Weather Sensor	Keeps track of solar radiation and pressure in an area.		(WSensorid, wid, Solar Radiation, Pressure)	Wsensorid	String	unique	wid
				Solar Radiation	INT	0 to 400 watt per square meter	
				Pressure	INT	870 to 1100 mbar	
Location	Which area temperature is being recorded from.	lid	(lid, wid, wSid, County, State)	lid	String	unique	wid, wStid
				County	String	Must be a valid location	
				State	String	Must be a valid location	
Historical Data	Details about weather in an area beforehand. Can be useful for making predictions or making interesting statistics.	hid	(hid, wid, Date, Rain Fall Information)	hid	String	unique	wid
				Date	INT	Must be a valid date MM/DD/YYYY	
				Rain Fall Information	INT	0 to 75 inches	
Location_County	Holds every county that a location is residing in.	County	(County, lid)	County	String	Must be a valid location	lid
has_specific	Relationship between Historical data and Location. Can be used to reference both.		(hid, lid)				hid, lid

5. Tools Used
    a. ERD Plus: https://erdplus.com/
    b. Draw.io: https://www.draw.io/