Homework: Relational Databases for an EOFS

Consider an Environmental Observation and Forecasting System [1]. Your task is design a relational schema to manage the *observational* data produced by an EOFS. In this assignment, we will not consider the forecasted data.

What to Turn In

1) *ER Diagram:* A pictorial representation of your schema that includes table names, attribute names, and relationships between tables. Does not have to follow any particular conventions, but clarity is important.
2) *Queries:* SQL queries exercising your schema that answer the questions below.
3) *Design Document:* Four or five paragraphs highlighting the important design decisions you made.

Sensor Data

At least two kinds of sensors are deployed at fixed stations in and around the Columbia River estuary: CTD instruments (Conductivity, Temperature, Depth) [2] and ADP instruments (Acoustic Doppler Profiler) [3].

CTD sensors report the **salinity** (a function of conductivity), **temperature**, and **elevation** (a function of pressure) at their position once every t seconds.

The ADP sensors emit sound waves vertically upward. The sound waves bounce off of particles in the water, indicating the velocity at which the particle is traveling. The result is a vertical profile of velocities, represented as an **array of velocity vectors** $(u, v)_d$ once every t seconds. Each value in the array corresponds to a particular depth d. The depths at which data are reported do not change from measurement to measurement.

Every hour, a telemetry system delivers one flat file per sensor. The CTD sensors report some metadata about the sensor and its configuration (serial number, model, sampling factor, calibration date), as well as a table of observed data (time, salinity, temperature, elevation)

Queries

Write the following queries over your schema. You may use any user-defined functions and data types you feel are appropriate, but make sure to describe them in detail. You may also decide to split a computation into two parts: a SQL statement to retrieve relevant data, and a client program to generate the final result. This strategy should not be strictly necessary, but you may employ it if you get stuck.

Q1) What is the distance between two stations red26 and tansy, ignoring the depth at which they are positioned?

Q2) Given two stations red26 and tansy equipped with CTD sensors, compute the difference in temperature at all times when both sensors are in fresh water? (Note: Fresh water has zero salinity.) A low number indicates that both sensors agree and are likely functioning properly.

Q3) For a given region defined by (xmin, xmax, ymin, ymax), report the maximum difference in fresh-water temperature reported at the same time by any two CTD sensors in the region. Basically, we hope to gather evidence that all sensors in a given region tend to agree on their temperature measurements whenever they agree that they are in fresh water. (Hint: Generalize your result from Q2, and then compute the maximum.)

Q4) Assume station red26 is equipped with both an ADP and a CTD. Report the time, the temperature, and the average velocity over depth for this station.

Q5) At station tansy, compute the change in velocity from time *t* to time *t*+1 for each depth and each time *t*. (Hint: the output should be a relation *Result(depth, time, delta_t)*)

References:

[1] "Research Challenges in Environmental Observation and Forecasting Systems" David Steere, Antonio Baptista, Dylan McNamee, Calton Pu, Jonathan Walpole, Proceedings of MOBICOMM 2000

[2] System Engineering & Fundamental Principles of CTD Accuracy SBE 911plus, Sea-Bird Electronics, http://www.seabird.com/products/spec_sheets/911princ.htm

[3] ADP Principles of Operation, SonTek/YSI, Inc., http://www.sontek.com/princop/adv/advpo.htm