Brett Fox

CS422

Project 1

Database Plan Overview

For Project 1, I plan on implementing a database that holds information on time series data uploaded by a Contributor and maintained by a Database Administrator (DBA). Attached to this document is a picture of my database model using [Crow's Foot notation](https://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model#Crow's_foot_notation).

Diagram

Description automatically generated

* The TS\_SET table contains information on any given set of time series data. A set of time series data is defined as a collection of time series data that shares the same unit (seconds, days, years, etc.), and starts at the same time value (September 7, 1999, for example). A set may only consist of one time series. In this case, it is a single-variable set. In the other case, it is a multivariable set. The table also contains the Boolean variable projection\_or\_not, which is 1 if the set is a projection and 0 if the data is not.
* The TIME SERIES table contains metadata about every specific time series. Every time series consists of a set of x-values that represent time and a series of y-values that represent another quantity. Each time series belongs to one (and only one) set. Though each time series in a given set shares the same time unit (x-axis in a time series), the series may have a different measurement unit (measured on the y-axis in a time series. Not time-related).
* The TS\_MEASUREMENT table contains the raw data for each time series. Data is not standardized using units in this table.
* The CONTRIBUTOR table consists of a list of contributors to a given time series set.
* The PAPER table consists of a list of papers referenced by a given time series set.
* The KEYWORD table consists of a list of keywords that appear in either the time series set, or a time series within a set.
* The DOMAIN table consists of the domain of a given time series set. A domain is defined as the unit of time used by the time series measurement (second, minute, year, etc.), and the first value of the time series.
* All other tables are “join” tables that allow many-to-many relationships between two of the tables listed above. They contain no additional unique information.

Some of the assumptions made by my database model:

* All specific time series belong to a set, even if it is only a set of one time series.
* Each time series appears in one and only one set.
* All time series in a set use the same domain (time measurement unit)
* Each set has one and only one domain.
* All information pertaining to the time series is to be stored in the database. The backend can make a local copy of it if it wishes.
* The TS\_MEASUREMENT metric stores data points for an individual time series in a unit-agnostic manner, only storing them as raw numbers for x and y values.

This database is an initial concept and any element of it is subject to change.