wrcc2hdb.py - Python based Climate Data Loader

# Overview

The wrcc2hdb.py Python data loader is designed to load monthly precipitation and average temperature from the Western Regional Climate Center (WRCC) to the HDB R tables. WRCC data is used to compute Upper Colorado consumptive use and loss (CUL) data, but could also be used for other purposes by mapping the loader to different time series. By default, data written by the loader includes an ‘O’ in the r\_base.overwrite\_flag field. Wrcc2hdb.py is located underneath the “python” subdirectory in the HDB repository and relies on the generic lib/hdb.py class, which provides database connection and access to HDB SQL packages like ts\_xfer for writing arrays of real data.

Graphical user interface, application

Description automatically generated

Figure : Organization of python loader directory.

# Installation/Requirements

The following python installation and packages are required:

* Python 3.9
* Cx\_oracle
* Pandas
* Argparse

# WRCC Sites and Data

WRCC sites are uniquely identified by 6-digit ID’s, with the first two digits representing a state. Each site also has a string name in WRCC’s system. Two monthly datatypes are currently supported by wrcc2hdb.py: precipitation and temperature. Precipitation is summed for the month, and temperature is averaged. Wrcc2hdb.py modifies 3 inputs to the WRCC URL query string to get a specific combination of site, datatype, and monthly summary, as shown in the example URL below. ‘020750’ is a site id, ‘avgt’ indicates average temperature (use ‘pcpn’ for precipitation), and ‘mave’ represents an average summary (use ‘msum’ for a sum).

[https://wrcc.dri.edu/WRCCWrappers.py?sodxtrmts+**020750**+por+por+**avgt**+none+**mave**+5+01+F](https://wrcc.dri.edu/WRCCWrappers.py?sodxtrmts+020750+por+por+avgt+none+mave+5+01+F)

The result of the example query string is a table of monthly average temperature at site 020750, also known as Betatakin, AZ. The record covers all available historical data for that site. In addition to each monthly value, a single character indicates how many days of missing data are associated with each month. ‘a’ means that a single day is missing, ‘b’ means two, and so on. A missing character indicates that a full month of data is present. This character will be written to the data\_flags field in r\_base on a per-data point basis. The screenshot below shows a portion of the table resulting from the example query string.

A picture containing text

Description automatically generated

Figure 2: A portion of the results table from the example WRCC query string.

# Instructions

## Site Datatype Mapping

WRCC sites and datatypes are mapped to HDB time series using the hdb\_ext\_data\_source and ref\_ext\_site\_data\_map tables. Wrcc2hdb.py requires a single external data source (primary key of the hdb\_ext\_data\_source table), which corresponds to a set of mappings in the ref\_ext\_site\_data\_map table. An example from the CUL project is shown below:



Figure : External data source entry in hdb\_ext\_data\_source table for CUL WRCC climate load.

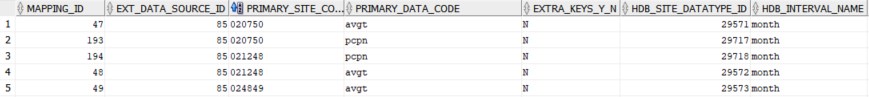


Figure : Portion of site datatype mappings in ref\_ext\_site\_data\_map table from the CUL WRCC climate external data source.

Note that only monthly data is supported by wrcc2hdb.py, so all entries in ref\_ext\_site\_data\_map should use ‘month’ for the hdb\_interval\_name field. Also, the loader will pull additional metadata that isn’t shown in the above screenshots for the r\_base table from the two mapping tables. Collection\_system\_id comes from hdb\_ext\_data\_source, while method\_id, computation\_id, and agen\_id come from ref\_ext\_site\_data\_map.

## Connection File

A connection file is used to supply database connection credentials. If stored in the hdb/python directory, use \*.auth to ensure that the file is ignored by Git. The screenshot below shows required connection file format.

Text

Description automatically generated

Figure : Database connection file format.

## Script Arguments

The loader can be run through the command line, either manually or through a scheduled process, with the following arguments. Additionally, “python wrcc2hdb.py -h” will display similar help information on the command line while using the loader.

* Required Arguments
  + -a/--authFile – path to database connection file
* Optional Arguments
  + -d/--dataSource - Name of data source to get time series mappings from. Defaults to “WRCC Load For CUL”
  + -m/--startMonth - Start month (m/yyyy) to begin loading data for. Default to January of the current year
  + -s/--site - 6-digit numeric ID of single WRCC site to load data for. Defaults to all sites matching the provided dataSource argument
* Optional Switches
  + --verbose - Print additional detail to command line as the process is executed
  + --nooverwrite - Do not write an ‘O’ to the r\_base.overwrite\_flag field